=> fil capl; d que 19; fil PASCAL, JICST-EPLUS, INSPEC, LIFESCI, BIOSIS, ANABSTR, SCISEARCH; d que 166; fil wpids; d que 13
FILE 'CAPLUS' ENTERED AT 16:02:11 ON 22 OCT 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

FILE COVERS 1907 - 22 Oct 2003 VOL 139 ISS 17
FILE LAST UPDATED: 21 Oct 2003 (20031021/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

L1	24	SEA	FILE=CAPLUS ABB=ON	BUSA W?/AU
L6	40840	SEA	FILE=CAPLUS ABB=ON	DATABASE#
L7	83251	SEA	FILE=CAPLUS ABB=ON	INFER?
L8	71908	SEA	FILE=CAPLUS ABB=ON	ALGORITHM#
L9	0	SEA	FILE=CAPLUS ABB=ON	L1 AND (L6 OR L7 OR L8)

FILE 'PASCAL' ENTERED AT 16:02:11 ON 22 OCT 2003

FILE 'JICST-EPLUS' ENTERED AT 16:02:11 ON 22 OCT 2003

FILE 'INSPEC' ENTERED AT 16:02:11 ON 22 OCT 2003
Compiled and produced by the IEE in association with FIZ KARLSRUHE

FILE 'LIFESCI' ENTERED AT 16:02:11 ON 22 OCT 2003

FILE 'BIOSIS' ENTERED AT 16:02:11 ON 22 OCT 2003

FILE 'ANABSTR' ENTERED AT 16:02:11 ON 22 OCT 2003

FILE 'SCISEARCH' ENTERED AT 16:02:11 ON 22 OCT 2003

L55	85	SEA		BUSA W?/AU OR BUSA, W?/AU
L56	142316	SEA		STRUCTUR?(3A) ACTIVIT?

```
L57      174626 SEA (MOLECUL? OR STRUCTUR?)(5A)(BIOLOGICAL? OR CHEMICAL?)
L58      610997 SEA (MOLECUL? OR STRUCTUR?)(5A)(PROCESS? OR FUNCTION? OR
                ACTIVIT?)
L59       98530 SEA INFERENCE#
L60       14988 SEA COOCCUR? OR CO OCCUR?
L66           0 SEA L55 AND (L56 OR L57 OR L58 OR L59 OR L60 OR L61 OR L62 OR
                L63 OR L64)


FILE 'WPIDS' ENTERED AT 16:02:12 ON 22 OCT 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE LAST UPDATED:          22 OCT 2003      <20031022/UP>
MOST RECENT DERWENT UPDATE:     200368        <200368/DW>
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE


>>> NEW WEEKLY SDI FREQUENCY AVAILABLE --> see NEWS  <<<

>>> PATENT IMAGES AVAILABLE FOR PRINT AND DISPLAY <<<

>>> FOR A COPY OF THE DERWENT WORLD PATENTS INDEX STN USER GUIDE,
    PLEASE VISIT:
 http://www.stn-international.de/training_center/patents/stn_guide.pdf <<<

>>> FOR DETAILS OF THE PATENTS COVERED IN CURRENT UPDATES, SEE
    http://thomsonderwent.com/coverage/latestupdates/           <<<

>>> FOR INFORMATION ON ALL DERWENT WORLD PATENTS INDEX USER
    GUIDES, PLEASE VISIT:
    http://thomsonderwent.com/support/userguides/               <<<


L3             3 SEA FILE=WPIDS ABB=ON  BUSA W?/AU


=> d ibib ab l3 1-3

L3   ANSWER 1 OF 3  WPIDS  COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER:     2002-454466 [48]   WPIDS
DOC. NO. CPI:         C2002-129179
TITLE:                Quantifying target gene expression in living cells that
                      possess a target gene of interest tagged with the binding
                      site for an RNA binding protein and fluorescently labeled
                      RNA binding polypeptide including an RNA binding domain.
DERWENT CLASS:        B04 D16
INVENTOR(S):          BUSA, W B
PATENT ASSIGNEE(S):   (CELL-N) CELLOMICS INC; (BUSA-I) BUSA W B
COUNTRY COUNT:        95
PATENT INFORMATION:

     PATENT NO   KIND DATE     WEEK      LA   PG
     -------------------------------------------
     WO 2002027031 A2 20020404 (200248)* EN   51
        RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ
            NL OA PT SD SE SL SZ TR TZ UG ZW
         W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK
            DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ
            LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD
            SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
     AU 2001094872 A  20020408 (200252)
     US 2003096243 A1 20030522 (200336)
```

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2002027031 A2		WO 2001-US30438	20010928
AU 2001094872 A		AU 2001-94872	20010928
US 2003096243 A1	Provisional	US 2000-236407P	20000928
		US 2001-965876	20010928

FILING DETAILS:

PATENT NO	KIND	PATENT NO
AU 2001094872 A	Based on	WO 2002027031

PRIORITY APPLN. INFO: US 2000-236407P 20000928; US 2001-965876 20010928

AB WO 200227031 A UPAB: 20020730

NOVELTY - Quantifying (M1) expression of target genes in living cells comprising:

(1) providing cells that possess a target gene of interest which has been tagged with the binding site for an RNA binding protein and a fluorescently labeled RNA binding polypeptide (I) that includes an RNA binding domain; and

(2) calculating the quantity of target gene expression in the cells using fluorescence signaling techniques.

DETAILED DESCRIPTION - Quantifying (M1) expression of one or more target genes in living cells comprising:

(a) providing cells that possess at least a first fluorescently labeled RNA binding polypeptide (I) which comprises first RNA binding domain (RBD1), and at least a first target gene of interest (T1) that has been modified to comprise one or more nucleic acid sequences encoding a first binding site (BS1) for RBD1 where, upon expression of (T1) into first target RNA, BS1 is specifically bound by the first fluorescently labeled (I);

(b) scanning the cells to obtain fluorescent signals from the first fluorescently labeled (I);

(c) determining fluorescent emission intensities from the first fluorescently labeled (I) at two different wavelengths;

(d) calculating a ratio of the fluorescent emission intensities from the first fluorescently labeled (I) at the two different wavelengths; and

(e) calculating a quantity of the first target RNA in the cells from the ratio.

An INDEPENDENT CLAIM is included for a fluorescently labeled (I) comprising:

(a) a non-naturally occurring amino acid sequence comprising:

(i) a nuclear export signal; and

(ii) an RNA binding domain; and

(b) a fluorophore pair such as a donor/acceptor pair for fluorescence resonance energy transfer (FRET), an excimer forming fluorophore pair, or an exciplex forming fluorophore pair.

USE - (M1) is useful for quantifying expression of one or more target genes in living cells which comprise two or more distinct populations of cells (claimed). The method is used to quantitate the expression of any target gene, including expression of protein-encoding messenger RNA genes, ribosomal RAN encoding genes, and transfer RNA encoding genes, so long as the RNA expression product from the target gene possesses a sequence or structure (the RNA tag) that is bound specifically by the RNA binding polypeptide being used.

Dwg.0/3

L3 ANSWER 2 OF 3 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER: 2001-496878 [54] WPIDS
CROSS REFERENCE: 2001-476263 [51]
DOC. NO. NON-CPI: N2001-368186
TITLE: Automated inference creation involves analyzing connection network constructed using records from inference database, to determine inference regarding physico-chemical relation between chemical or biological molecules.
DERWENT CLASS: T01
INVENTOR(S): **BUSA, W B**
PATENT ASSIGNEE(S): (CELL-N) CELLOMICS INC
COUNTRY COUNT: 95
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2001055950	A2	20010802	(200154)*	EN	38

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

AU 2001032928 A 20010807 (200174)
EP 1252596 A2 20021030 (200279) EN
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2001055950	A2	WO 2001-US2245	20010124
AU 2001032928	A	AU 2001-32928	20010124
EP 1252596	A2	EP 2001-905006	20010124
		WO 2001-US2245	20010124

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2001032928	A	Based on	WO 2001055950
EP 1252596	A2	Based on	WO 2001055950

PRIORITY APPLN. INFO: US 2001-769169 20010124; US 2000-177964P 20000125

AB WO 200155950 A UPAB: 20021209
NOVELTY - Co-occurrence count is set to starting values of co-occurring preset name and filtered chemical or biological molecule name, when filtered name is not stored in inference database (24,26). Co-occurrence count is incremented for each pair of preset name, when stored in database. Connection network is constructed using records from database and analyzed to determine inferences regarding relationships between chemical or biological molecules.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(a) Method for checking automatically created inferences;

(b) Automated inference system

USE - For creating automated inferences for physico-chemical interactions through co-occurrence analysis of inference databases.

ADVANTAGE - Allows scientists and researchers to automatically create and check inferences of physico-chemical interaction through co-occurrence analysis of indexed databases. Facilitates user's understanding of

biological functions, such as cell function, to design experiments more intelligently and to analyze experimental results more thoroughly. Helps drug discovery scientists select better targets for pharmaceutical intervention in hope of curing diseases.

DESCRIPTION OF DRAWING(S) - The figure shows the exemplary experimental data storage system for storing experimental data.

Inference database 24,26

Dwg.1/4

L3 ANSWER 3 OF 3 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER: 2001-476263 [51] WPIDS
CROSS REFERENCE: 2001-496878 [54]
DOC. NO. NON-CPI: N2001-352481
DOC. NO. CPI: C2001-142902
TITLE: Strength measurement of co-occurrence data for automated interference of physico-chemical interaction knowledge, involves determining if co-occurrence between at least two chemical or biological molecule names is non-trivial.
DERWENT CLASS: B04 D16 T01
INVENTOR(S): **BUSA, W B**
PATENT ASSIGNEE(S): (CELL-N) CELLOMICS INC; (BUSA-I) BUSA W B
COUNTRY COUNT: 95
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2001055951	A2	20010802	(200151)*	EN	63

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

AU 2001029744	A	20010807	(200174)		
AU 2001032928	A	20010807	(200174)		
US 2002002559	A1	20020103	(200207)		
US 2002004792	A1	20020110	(200208)		
EP 1252598	A2	20021030	(200279)	EN	

R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
WO 2001055951	A2		WO 2001-US2294	20010124
AU 2001029744	A		AU 2001-29744	20010124
AU 2001032928	A		AU 2001-32928	20010124
US 2002002559	A1	Provisional	US 2000-177964P	20000125
			US 2001-769169	20010124
US 2002004792	A1	Provisional	US 2000-177964P	20000125
		Provisional	US 2000-201105P	20000502
			US 2001-768686	20010124
EP 1252598	A2		EP 2001-946969	20010124
			WO 2001-US2294	20010124

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2001029744	A	Based on	WO 2001055951
AU 2001032928	A	Based on	WO 2001055950
EP 1252598	A2	Based on	WO 2001055951

PRIORITY APPLN. INFO: US 2001-768686 20010124; US 2000-177964P 20000125; US 2000-201105P 20000502; US 2001-769169 20010124

AB WO 200155951 A UPAB: 20021209

NOVELTY - A strength of co-occurrence data is measured by extracting at least two chemical or biological molecule names from database record; and determining likelihood statistic for co-occurrence reflecting physico-chemical interactions between the two molecule names, and applying it to the co-occurrence to determine if co-occurrence between the molecule names is non-trivial.

DETAILED DESCRIPTION - Strength measurement of co-occurrence data involves extracting at least two chemical or biological molecule names from database record from an interference database; determining likelihood statistic for co-occurrence reflecting physico-chemical interactions between the two molecule names (A and B); and applying the likelihood statistic to the co-occurrence to determine if the co-occurrence between molecule A and molecule B is non-trivial. The interference database includes those records created from an indexed literature database. The two molecule names co-occur in at least one record in an indexed scientific literature database.

An INDEPENDENT CLAIM is also included for:

(1) a method of contextual querying of co-occurrence data comprising selecting a target node from a first list of nodes connected by arcs in a connection network; creating a second list of nodes by considering other nodes that are neighbors of the target node and other nodes in prior to the target node in the connection network; selecting a next node from the second list of nodes using the co-occurrence values, in which the next node is next after the target node in the pre-determined order for the connection network based on the co-occurrence values;

(2) method of query polling of co-occurrence data comprising selecting a position in connection network for an unknown target node from a first list of nodes; determining a second list of nodes prior to the position of unknown target node in the connection network; determining a third list of nodes subsequent to the position of unknown target node in the connection network; determining a fourth list of nodes included in both the second and the third lists of nodes; and determining an identity for the unknown target node by selecting a node from the fourth list of nodes using likelihood statistic; and

(3) a method for creating automated biological interferences comprising constructing a connection network using at least one database record from an interference database; applying likelihood statistics analysis methods to the connection network; generating automatically at least one biological interferences relationships between chemical or biological molecules or biological processes using the results from the likelihood statistic analysis methods.

USE - The method is for automated interference of physico-chemical interaction knowledge from databases of term co-occurrence data. It can also be used to facilitate a user's understanding of biological functions, e.g. cell functions, to design experiments, and to analyze experiment results.

ADVANTAGE - The method helps drug discovery scientists select better targets for pharmaceutical intervention of curing diseases. It may also help facilitate the abstraction of knowledge from information for biological experimental data and provides new bioinformatic techniques.

Dwg.0/9

```
=> fil capl; d que l17; d que l19; d que l24; d que l33; d que l34
FILE 'CAPLUS' ENTERED AT 16:04:03 ON 22 OCT 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)
```



```
FILE COVERS 1907 - 22 Oct 2003  VOL 139 ISS 17
FILE LAST UPDATED: 21 Oct 2003  (20031021/ED)
```

This file contains CAS Registry Numbers for easy and accurate substance identification.

```
L6          40840 SEA FILE=CAPLUS ABB=ON  DATABASE#
L7          83251 SEA FILE=CAPLUS ABB=ON  INFER?
L8          71908 SEA FILE=CAPLUS ABB=ON  ALGORITHM#
L10        213037 SEA FILE=CAPLUS ABB=ON  (CHEMICAL OR BIOLOGICAL)(3A)(STRUCTUR?
                  OR MOLECUL?)
L12         55302 SEA FILE=CAPLUS ABB=ON  L10 AND (PROCESS? OR FUNCTION?)
L13           362 SEA FILE=CAPLUS ABB=ON  L12 AND L7
L14            20 SEA FILE=CAPLUS ABB=ON  (L6 OR L8) AND L13
L16        239459 SEA FILE=CAPLUS ABB=ON  STATISTIC?
L17             4 SEA FILE=CAPLUS ABB=ON  L16 AND L14


L6          40840 SEA FILE=CAPLUS ABB=ON  DATABASE#
L7          83251 SEA FILE=CAPLUS ABB=ON  INFER?
L10        213037 SEA FILE=CAPLUS ABB=ON  (CHEMICAL OR BIOLOGICAL)(3A)(STRUCTUR?
                  OR MOLECUL?)
L18            46 SEA FILE=CAPLUS ABB=ON  L6(3A)L7
L19             3 SEA FILE=CAPLUS ABB=ON  L10 AND L18


L6          40840 SEA FILE=CAPLUS ABB=ON  DATABASE#
L7          83251 SEA FILE=CAPLUS ABB=ON  INFER?
L8          71908 SEA FILE=CAPLUS ABB=ON  ALGORITHM#
L10        213037 SEA FILE=CAPLUS ABB=ON  (CHEMICAL OR BIOLOGICAL)(3A)(STRUCTUR?
                  OR MOLECUL?)
L20         11937 SEA FILE=CAPLUS ABB=ON  L10(10A)(PROCESS? OR FUNCTION?)
L21            52 SEA FILE=CAPLUS ABB=ON  L20 AND L7
L23             6 SEA FILE=CAPLUS ABB=ON  L21 AND (L6 OR L8)
L24             1 SEA FILE=CAPLUS ABB=ON  GENETIC/TI AND L23


L6          40840 SEA FILE=CAPLUS ABB=ON  DATABASE#
L8          71908 SEA FILE=CAPLUS ABB=ON  ALGORITHM#
L25          1971 SEA FILE=CAPLUS ABB=ON  COOCCUR? OR CO OCCUR?
```

```
L31       60160 SEA FILE=CAPLUS ABB=ON  MOLECULAR STRUCTURE-BIOLOGICAL
                ACTIVITY RELATIONSHIP/CT
L33           1 SEA FILE=CAPLUS ABB=ON  L25 AND L31 AND (L6 OR L8)



L6        40840 SEA FILE=CAPLUS ABB=ON  DATABASE#
L7        83251 SEA FILE=CAPLUS ABB=ON  INFER?
L8        71908 SEA FILE=CAPLUS ABB=ON  ALGORITHM#
L29        6776 SEA FILE=CAPLUS ABB=ON  BIOINFORMATIC#
L31       60160 SEA FILE=CAPLUS ABB=ON  MOLECULAR STRUCTURE-BIOLOGICAL
                ACTIVITY RELATIONSHIP/CT
L34           2 SEA FILE=CAPLUS ABB=ON  L7 AND L31 AND (L6 OR L8 OR L29)


=> s l17 or l19 or l24 or l33 or l34

L97           9 L17 OR L19 OR L24 OR L33 OR L34

=> fil wpids; d que 142; d que 144; d que 150;d que 154

FILE 'WPIDS' ENTERED AT 16:04:04 ON 22 OCT 2003
COPYRIGHT (C) 2003 THOMSON DERWENT

FILE LAST UPDATED:          22 OCT 2003      <20031022/UP>
MOST RECENT DERWENT UPDATE:     200368        <200368/DW>
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE


>>> NEW WEEKLY SDI FREQUENCY AVAILABLE --> see NEWS  <<<

>>> PATENT IMAGES AVAILABLE FOR PRINT AND DISPLAY <<<

>>> FOR A COPY OF THE DERWENT WORLD PATENTS INDEX STN USER GUIDE,
    PLEASE VISIT:
 http://www.stn-international.de/training_center/patents/stn_guide.pdf <<<

>>> FOR DETAILS OF THE PATENTS COVERED IN CURRENT UPDATES, SEE
    http://thomsonderwent.com/coverage/latestupdates/           <<<

>>> FOR INFORMATION ON ALL DERWENT WORLD PATENTS INDEX USER
    GUIDES, PLEASE VISIT:
    http://thomsonderwent.com/support/userguides/               <<<


L36       14800 SEA FILE=WPIDS ABB=ON  INFER?
L37       78557 SEA FILE=WPIDS ABB=ON  DATABASE# OR ALGORITHM#
L40        5986 SEA FILE=WPIDS ABB=ON  (CHEMICAL OR BIOLOGICAL)(3A)(MOLECULE#
                OR STRUCTUR?)
L41           6 SEA FILE=WPIDS ABB=ON  L40 AND L36 AND L37
L42           3 SEA FILE=WPIDS ABB=ON  L41 NOT INFERT?



L36       14800 SEA FILE=WPIDS ABB=ON  INFER?
L37       78557 SEA FILE=WPIDS ABB=ON  DATABASE# OR ALGORITHM#
L43          46 SEA FILE=WPIDS ABB=ON  L36(3A)L37
L44           4 SEA FILE=WPIDS ABB=ON  (CHEMICAL# OR CHEMISTRY OR BIOLOGICAL)
                AND L43



L37       78557 SEA FILE=WPIDS ABB=ON  DATABASE# OR ALGORITHM#
```

```
L45          545 SEA FILE=WPIDS ABB=ON  STRUCTUR?(2A)ACTIVIT?
L47           36 SEA FILE=WPIDS ABB=ON  L37 AND L45
L48           24 SEA FILE=WPIDS ABB=ON  L47 AND T/DC
L49           13 SEA FILE=WPIDS ABB=ON  L48 AND B04/DC
L50            4 SEA FILE=WPIDS ABB=ON  L49 AND (ACTIVITY OR DESCRIPTOR#)/TI


L37        78557 SEA FILE=WPIDS ABB=ON  DATABASE# OR ALGORITHM#
L40         5986 SEA FILE=WPIDS ABB=ON  (CHEMICAL OR BIOLOGICAL)(3A)(MOLECULE#
                 OR STRUCTUR?)
L45          545 SEA FILE=WPIDS ABB=ON  STRUCTUR?(2A)ACTIVIT?
L51          369 SEA FILE=WPIDS ABB=ON  CONNECTION NETWORK#
L52       160517 SEA FILE=WPIDS ABB=ON  NODE# OR ARC#
L54            3 SEA FILE=WPIDS ABB=ON  (L40 OR L45) AND (L51 OR L52) AND L37
                 AND (INTERACTION# OR RELATION?)/TI


=> s (l42 or l44 or l50 or l54) not l3

L98            9 (L42 OR L44 OR L50 OR L54) NOT L3

=> fil PASCAL, JICST-EPLUS, INSPEC, LIFESCI, BIOSIS, ANABSTR, SCISEARCH

FILE 'PASCAL' ENTERED AT 16:04:06 ON 22 OCT 2003
Any reproduction or dissemination in part or in full,
by means of any process and on any support whatsoever
is prohibited without the prior written agreement of INIST-CNRS.
COPYRIGHT (C) 2003 INIST-CNRS. All rights reserved.

FILE 'JICST-EPLUS' ENTERED AT 16:04:06 ON 22 OCT 2003
COPYRIGHT (C) 2003 Japan Science and Technology Corporation (JST)

FILE 'INSPEC' ENTERED AT 16:04:06 ON 22 OCT 2003
Compiled and produced by the IEE in association with FIZ KARLSRUHE
COPYRIGHT 2003 (c) INSTITUTION OF ELECTRICAL ENGINEERS (IEE)

FILE 'LIFESCI' ENTERED AT 16:04:06 ON 22 OCT 2003
COPYRIGHT (C) 2003 Cambridge Scientific Abstracts (CSA)


FILE 'BIOSIS' ENTERED AT 16:04:06 ON 22 OCT 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'ANABSTR' ENTERED AT 16:04:06 ON 22 OCT 2003
COPYRIGHT (c) 2003 THE ROYAL SOCIETY OF CHEMISTRY (RSC)

FILE 'SCISEARCH' ENTERED AT 16:04:06 ON 22 OCT 2003
COPYRIGHT 2003 THOMSON ISI

=> d que l71; d que l68;d que l75; d que l80 ;d que l87; d que l83; d que l88; d que l90;
d que l96

L59        98530 SEA INFERENCE#
L61       153155 SEA PHYSICOCHEMICAL OR PHYSICO CHEMICAL
L64      1326996 SEA DATABASE# OR ALGORITHM#
L67         2346 SEA L59(3A) L64
L71            0 SEA L67 AND L61


L57       174626 SEA (MOLECUL? OR STRUCTUR?)(5A)(BIOLOGICAL? OR CHEMICAL?)
L59        98530 SEA INFERENCE#
```

```
L64      1326996 SEA DATABASE# OR ALGORITHM#
L67         2346 SEA [illegible] L64

  L56      142316 SEA ABB=ON  STRUCTUR?(3A) ACTIVI
  L57      174626 SEA ABB=ON  (MOLECUL? OR STRUCTUI
                  CHEMICAL?)
  L58      610997 SEA ABB=ON  (MOLECUL? OR STRUCTUR?)(
                  OR ACTIVIT?)
  L60       14988 SEA ABB=ON  COOCCUR? OR CO OCCUR?
  L61      153155 SEA ABB=ON  PHYSICOCHEMICAL OR PHYSICO CHEMICAL
  L62         363 SEA ABB=ON  CONNECTION NETWORK#
  L63      596298 SEA ABB=ON  NODE# OR NODAL? OR ARC#
  L64     1326996 SEA ABB=ON  DATABASE# OR ALGORITHM#

L78           16 SEA ((L56 OR L57 OR L58) OR L61) AND (L62 OR L63 OR L64) AND
                 L60
L80            2 SEA L78 AND PROTEIN/TI



L56       142316 SEA STRUCTUR?(3A) ACTIVIT?
L57       174626 SEA (MOLECUL? OR STRUCTUR?)(5A)(BIOLOGICAL? OR CHEMICAL?)
L58       610997 SEA (MOLECUL? OR STRUCTUR?)(5A)(PROCESS? OR FUNCTION? OR
                 ACTIVIT?)
L59        98530 SEA INFERENCE#
L61       153155 SEA PHYSICOCHEMICAL OR PHYSICO CHEMICAL
L82        11219 SEA L61 AND (L56 OR L57 OR L58)
L86           12 SEA L82 AND L59
L87            1 SEA L86 AND REASONING/TI



L56       142316 SEA STRUCTUR?(3A) ACTIVIT?
L57       174626 SEA (MOLECUL? OR STRUCTUR?)(5A)(BIOLOGICAL? OR CHEMICAL?)
L58       610997 SEA (MOLECUL? OR STRUCTUR?)(5A)(PROCESS? OR FUNCTION? OR
                 ACTIVIT?)
L59        98530 SEA INFERENCE#
L61       153155 SEA PHYSICOCHEMICAL OR PHYSICO CHEMICAL
L62          363 SEA CONNECTION NETWORK#
L63       596298 SEA NODE# OR NODAL? OR ARC#
L64      1326996 SEA DATABASE# OR ALGORITHM#
L82        11219 SEA L61 AND (L56 OR L57 OR L58)
L83            1 SEA (L62 OR L63 OR L64) AND L59 AND L82



L56       142316 SEA STRUCTUR?(3A) ACTIVIT?
L57       174626 SEA (MOLECUL? OR STRUCTUR?)(5A)(BIOLOGICAL? OR CHEMICAL?)
L58       610997 SEA (MOLECUL? OR STRUCTUR?)(5A)(PROCESS? OR FUNCTION? OR
                 ACTIVIT?)
L61       153155 SEA PHYSICOCHEMICAL OR PHYSICO CHEMICAL
L62          363 SEA CONNECTION NETWORK#
L63       596298 SEA NODE# OR NODAL? OR ARC#
L64      1326996 SEA DATABASE# OR ALGORITHM#
L82        11219 SEA L61 AND (L56 OR L57 OR L58)
L88            1 SEA L82 AND L64 AND (L62 OR L63)



L56       142316 SEA STRUCTUR?(3A) ACTIVIT?
L57       174626 SEA (MOLECUL? OR STRUCTUR?)(5A)(BIOLOGICAL? OR CHEMICAL?)
L58       610997 SEA (MOLECUL? OR STRUCTUR?)(5A)(PROCESS? OR FUNCTION? OR
                 ACTIVIT?)
```

```
L59       98530 SEA INFERENCE#
L60       14988 SEA COOCCUR? OR CO OCCUR?
L61      153155 SEA PHYSICOCHEMICAL OR PHYSICO CHEMICAL
L64     1326996 SEA DATABASE# OR ALGORITHM#
L82       11219 SEA L61 AND (L56 OR L57 OR L58)
L89         300 SEA L82 AND L64
L90           1 SEA (L59 OR L60) AND L89



L56      142316 SEA STRUCTUR?(3A) ACTIVIT?
L57      174626 SEA (MOLECUL? OR STRUCTUR?)(5A)(BIOLOGICAL? OR CHEMICAL?)
L58      610997 SEA (MOLECUL? OR STRUCTUR?)(5A)(PROCESS? OR FUNCTION? OR
                ACTIVIT?)
L61      153155 SEA PHYSICOCHEMICAL OR PHYSICO CHEMICAL
L64     1326996 SEA DATABASE# OR ALGORITHM#
L82       11219 SEA L61 AND (L56 OR L57 OR L58)
L89         300 SEA L82 AND L64
L91         242 SEA L89 AND (CHEMICAL OR CHEMISTRY OR BIOLOG?)
L96          29 SEA L91 AND (RELATIONAL? OR NON SEQUENCE OR PRO OR BANK OR
                FOLD OR PHYSEAN OR DESCRIPTOR#)/TI


=> s l68 or l75 or l80 or l87 or l83 or l88 or l90 or l96

L99          36 L68 OR L75 OR L80 OR L87 OR L83 OR L88 OR L90 OR L96

=> dup rem l97,l99,l98
FILE 'CAPLUS' ENTERED AT 16:04:25 ON 22 OCT 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'PASCAL' ENTERED AT 16:04:25 ON 22 OCT 2003
Any reproduction or dissemination in part or in full,
by means of any process and on any support whatsoever
is prohibited without the prior written agreement of INIST-CNRS.
COPYRIGHT (C) 2003 INIST-CNRS. All rights reserved.

FILE 'INSPEC' ENTERED AT 16:04:25 ON 22 OCT 2003
Compiled and produced by the IEE in association with FIZ KARLSRUHE
COPYRIGHT 2003 (c) INSTITUTION OF ELECTRICAL ENGINEERS (IEE)

FILE 'LIFESCI' ENTERED AT 16:04:25 ON 22 OCT 2003
COPYRIGHT (C) 2003 Cambridge Scientific Abstracts (CSA)


FILE 'BIOSIS' ENTERED AT 16:04:25 ON 22 OCT 2003
COPYRIGHT (C) 2003 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'SCISEARCH' ENTERED AT 16:04:25 ON 22 OCT 2003
COPYRIGHT 2003 THOMSON ISI

FILE 'WPIDS' ENTERED AT 16:04:25 ON 22 OCT 2003
COPYRIGHT (C) 2003 THOMSON DERWENT
PROCESSING COMPLETED FOR L97
PROCESSING COMPLETED FOR L99
PROCESSING COMPLETED FOR L98
L100         45 DUP REM L97 L99 L98 (9 DUPLICATES REMOVED)
                ANSWERS '1-9' FROM FILE CAPLUS
                ANSWERS '10-16' FROM FILE PASCAL
                ANSWERS '17-20' FROM FILE INSPEC
                ANSWERS '21-22' FROM FILE LIFESCI
```

ANSWERS '23-25' FROM FILE BIOSIS
ANSWERS '26-36' FROM FILE SCISEARCH
ANSWERS '37-45' FROM FILE WPIDS

=> d ibib ab 1-45; fil hom

L100 ANSWER 1 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2002:329153 CAPLUS
DOCUMENT NUMBER: 137:59817
TITLE: Structural similarity to link sequence space: new potential superfamilies and implications for structural genomics
AUTHOR(S): Aloy, Patrick; Oliva, Baldomero; Querol, Enrique; Aviles, Francesc X.; Russell, Robert B.
CORPORATE SOURCE: EMBL, Heidelberg, D-69117, Germany
SOURCE: Protein Science (2002), 11(5), 1101-1116
CODEN: PRCIEI; ISSN: 0961-8368
PUBLISHER: Cold Spring Harbor Laboratory Press
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The current pace of **structural biol.** now means that protein three-dimensional structure can be known before protein **function**, making methods for assigning homol. via structure comparison of growing importance. Previous research has suggested that sequence similarity after structure-based alignment is one of the best discriminators of homol. and often **functional** similarity. Here, we exploit this observation, together with a merger of protein structure and sequence **databases**, to predict distant homologous relationships. We use the Structural Classification of Proteins (SCOP) **database** to link sequence alignments from the SMART and Pfam **databases**. We thus provide new alignments that could not be constructed easily in the absence of known three-dimensional structures. We then extend the method of Murzin (1993b) to assign **statistical** significance to sequence identities found after structural alignment and thus suggest the best link between diverse sequence families. We find that several distantly related protein sequence families can be linked with confidence, showing the approach to be a means for **inferring** homologous relationships and thus possible **functions** when proteins are of known structure but of unknown **function**. The anal. also finds several new potential superfamilies, where inspection of the assocd. alignments and superimpositions reveals conservation of unusual structural features or co-location of conserved amino acids and bound substrates. We discuss implications for Structural Genomics initiatives and for improvements to sequence comparison methods.
REFERENCE COUNT: 58 THERE ARE 58 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L100 ANSWER 2 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2001:477182 CAPLUS
DOCUMENT NUMBER: 136:161855
TITLE: A model for phylogenetic **inference** using **structural** and **chemical** covariates
AUTHOR(S): Tavare, Simon; Adams, Dean C.; Fedrigo, Olivier; Naylor, Gavin J. P.
CORPORATE SOURCE: Departments of Biological Sciences, Mathematics and Preventative Medicine, University of Southern California, Los Angeles, CA, 90089, USA
SOURCE: Pacific Symposium on Biocomputing 2001, Mauna Lani, HI, United States, Jan. 3-7, 2001 (2001), 215-225. Editor(s): Altman, Russ B. World Scientific Publishing Co. Pte. Ltd.: Singapore, Singapore.
CODEN: 69BLFC
DOCUMENT TYPE: Conference

LANGUAGE: English
AB We investigated whether or not evolutionary change in DNA sequence data was homogeneous across different classes of base pairs. DNA sequences for eight protein-coding mitochondrial genes were obtained for 38 vertebrate taxa from GenBank. Each nucleotide site in the alignment was classified according to a no. of covariates, including its codon position, genetic code degeneracy, and hydrophobicity. The evolutionary transition matrix for each base was estd. by tracing implied character changes under parsimony on a known phylogenetic tree. Canonical variates analyses of the **inferred** transition matrixes were performed for each gene to det. whether or not different classes of bases behaved similarly. We found five distinct clusters of transition matrixes that could be roughly defined by combinations of codon position and degeneracy. This pattern was consistent among all genes. A stochastic model of rate variation based on the interaction of the covariates was developed to assess the **statistical** significance of the clusters. The five-group classification was found to explain significantly more sequence variation than did a codon only classification, a codon degeneracy classification, or a codon and degeneracy classification. The same five-group classification was found for all genes tested, suggesting a common **process** underlying the mol. evolution of the mitochondrial genome. These results confirm that there are classes of base pairs that evolve differently, and suggest that models of sequence evolution that incorporate covariate information may be useful in developing nucleotide substitution models that more accurately reflect evolutionary history.
REFERENCE COUNT: 24 THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L100 ANSWER 3 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2000:885510 CAPLUS
DOCUMENT NUMBER: 135:205926
TITLE: **Genetic** network **inference**: from co-expression clustering to reverse engineering
AUTHOR(S): D'haeseleer, Patrik; Liang, Shoudan; Somogyi, Roland
CORPORATE SOURCE: Department of Computer Science, University of New Mexico, Albuquerque, NM, 87131, USA
SOURCE: Bioinformatics (2000), 16(8), 707-726
CODEN: BOINFP; ISSN: 1367-4803
PUBLISHER: Oxford University Press
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 103 refs. Motivation: Advances in mol. biol., anal. and computational technologies are enabling us to systematically investigate the complex **mol. processes** underlying **biol.** systems. In particular, using high-throughput gene expression assays, we are able to measure the output of the gene regulatory network. We aim here to review datamining and modeling approaches for conceptualizing and unraveling the functional relationships implicit in these datasets. Clustering of co-expression profiles allows us to **infer** shared regulatory inputs and functional pathways. We discuss various aspects of clustering, ranging from distances measures to clustering **algorithms** and multiple-cluster memberships. More advanced anal. aims to **infer** causal connections between genes directly, i.e. who is regulating whom and how. We discuss several approaches to the problem of reverse engineering of genetic networks, from discrete Boolean networks, to continuous linear and non-linear models. We conclude that the combination of predictive modeling with systematic exptl. verification will be required to gain a deeper insight into living organisms, therapeutic targeting and bioengineering.
REFERENCE COUNT: 103 THERE ARE 103 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L100 ANSWER 4 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2000:841321 CAPLUS
DOCUMENT NUMBER: 134:112480
TITLE: 13C NMR chemical shifts can predict disulfide bond formation
AUTHOR(S): Sharma, Deepak; Rajarathnam, Krishna
CORPORATE SOURCE: Department of Human Biological Chemistry and Genetics and Sealy Center for Structural Biology, University of Texas Medical Branch, Galveston, TX, 77555-1055, USA
SOURCE: Journal of Biomolecular NMR (2000), 18(2), 165-171
CODEN: JBNME9; ISSN: 0925-2738
PUBLISHER: Kluwer Academic Publishers
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The presence of disulfide bonds can be detected unambiguously only by x-ray crystallog., and otherwise must be **inferred** by chem. methods. In this study we demonstrate that 13C NMR chem. shifts are diagnostic of disulfide bond formation, and can discriminate between cysteine in the reduced (free) and oxidized (disulfide bonded) state. A **database** of cysteine 13CC.alpha. and C.beta. chem. shifts was constructed from the BioMagResBank (BMRB) and Sheffield **databases**, and published journals. **Statistical** anal. indicated that the C.beta. shift is extremely sensitive to the redox state, and can predict the disulfide-bonded state. Further, chem. shifts in both states occupy distinct clusters as a **function** of secondary structure in the C.alpha./C.beta. chem. shift map. On the basis of these results, we provide simple ground rules for predicting the redox state of cysteines; these rules could be used effectively in NMR structure detn., predicting new folds, and in protein folding studies.

REFERENCE COUNT: 8 THERE ARE 8 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L100 ANSWER 5 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1996:151829 CAPLUS
DOCUMENT NUMBER: 124:242855
TITLE: A comparison of some commercially available structural descriptors and clustering **algorithms**
AUTHOR(S): Brown, Robert D.; Bures, Mark G.; Martin, Yvonne C.
CORPORATE SOURCE: Pharm. Prod. Div., Abbott Lab., Abbott Part, IL, 60064, USA
SOURCE: Proceedings of the First Electronic Computational Chemistry Conference [CD-ROM] (1995), Meeting Date 1994, Paper 12. Editor(s): Bachrach, Steven M. ARInternet Corp.: Landover, Md.
CODEN: 62MDAN
DOCUMENT TYPE: Conference
LANGUAGE: English

AB Clustering methods play an important part in the selection of compds. from chem. **databases** for both purchase and biol. screening. These clustering methods usually rely on descriptors which encode the structural features of the mols. in the **databases**. Structural descriptors allow the similarities of pairs of mols. to be calcd. from the **co-occurrence** of these features. Clusters may then be assembled on the basis of the similarity measures. A no. of methods exist within com. available **database** searching software to produce these descriptors. In this paper the relative merits of some of these descriptors, which variously describe the two-dimensional and three-dimensional content of mols., are examd. Two com. available clustering **algorithms** are also compared, one hierarchical and one non-hierarchical. All comparisons are based on the ability of the methods to produce sets of clusters in which biol. active and inactive structures do not occur in the same clusters. The various descriptors of two-dimensional structure perform better in this respect, particularly

when used in combination with the hierarchical clustering method.

L100 ANSWER 6 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1993:670093 CAPLUS
DOCUMENT NUMBER: 119:270093
TITLE: Similarity criteria for **chemical structures** and reactions
AUTHOR(S): Gasteiger, Johann; Ihlenfeldt, Wolf D.
CORPORATE SOURCE: Inst. Org. Chem., Tech. Univ. Munich, Garching, W-8046, Germany
SOURCE: Chem. Struct. 2 Proc. Int. Conf., 2nd (1993), Meeting Date 1990, 423-38. Editor(s): Warr, Wendy A. Springer: Berlin, Germany. CODEN: 59IUAO
DOCUMENT TYPE: Conference
LANGUAGE: English
AB New definitions of similarity of **chem. structures** are presented that are based on finding building blocks for synthesis and on general types of reactions. The merits of these similarity criteria in analyzing a **database** of structures and in designing org. syntheses are illustrated. Reaction similarities are based on values for electronic and energy effects. They allow novel search strategies for reaction **databases** and **inferences** on reaction conditions.

L100 ANSWER 7 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1991:631325 CAPLUS
DOCUMENT NUMBER: 115:231325
TITLE: A combined model of multi-resonance subspectra/substructure and DARC topological structure representation. Local and global knowledge in the carbon-13 NMR DARC **database**
AUTHOR(S): Carabedian, Michel; Dubois, Jacques Emile
CORPORATE SOURCE: Inst. Topol. Dyn. Syst., Univ. Paris 7, Paris, 75005, Fr.
SOURCE: Journal of Chemical Information and Computer Sciences (1991), 31(4), 564-74 CODEN: JCISD8; ISSN: 0095-2338
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The structural and spectral information in a 13C NMR **database** can be represented by means of a model which relates substructural fragments to subspectral features for multiple resonances. The substructural part of this model contains a concise DARC description of the structural part with a partially generic ELCOb which is assocd. with all the spectral information pertaining to the focal atom (Fo) and its neighboring carbons (Ai). In the spectral information, the concentric environmental view is shifted from the focal atom to the neighboring positions. This leads to overlap in the views and redundancy in the information and a dissym. phys. perception which formally, is broader than the substructural view. New substructural subspectral local and global knowledge **functions** of this model are managed with holog. techniques. Formalized local and global knowledge is described **statistically** by juxtaposition of the .delta.13CFo .times. .delta.13CAi correlation plane supporting the 3-dimensional occurrence distributions. Use of the **inferential** ability of these planes is facilitated by a table which correlates the repartitioning of the .sigma.- and .pi.-bonds in Fo-Ai atom pairs.

L100 ANSWER 8 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1989:624827 CAPLUS
DOCUMENT NUMBER: 111:224827
TITLE: Searching for pharmacophores in large coordinate data

bases and its use in drug design
AUTHOR(S): Sheridan, Robert P.; Rusinko, Andrew, III; Nilakantan, Ramaswamy; Venkataraghavan, R.
CORPORATE SOURCE: Med. Res. Div., American Cyanamid, Pearl River, NY, 10965, USA
SOURCE: Proceedings of the National Academy of Sciences of the United States of America (1989), 86(20), 8165-9
CODEN: PNASA6; ISSN: 0027-8424
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Pharmacophores, 3-dimensional arrangements of chem. groups essential for biol. activity, are being proposed in increasing nos. The authors developed a system to search data bases of 3-dimensional coordinates for compds. that contain a particular pharmacophore. The coordinates can be derived from expt. (e.g., Cambridge Crystal **Database**) or be generated from data bases of connection tables (e.g., Cyanamid Labs. proprietary compds.) via the program CONCORD. The authors discuss the results of searches for 3 sample pharmacophores. Two have been proposed by others based on the conformational anal. of active compds., and one is **inferred** from the crystal structure of a protein-ligand complex. These examples show that such searches can identify classes of compds. that are structurally different from the compds. from which the pharmacophore was derived but are known to have the appropriate biol. activity. Occasionally, the searches find bond "frameworks" in which the important groups are rigidly held in the proper geometry. These may suggest new structural classes for synthesis.

L100 ANSWER 9 OF 45 CAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1985:470570 CAPLUS
DOCUMENT NUMBER: 103:70570
TITLE: DARC system for documentation and artificial intelligence in chemistry
AUTHOR(S): Dubois, Jacques Emile; Sobel, Yves
CORPORATE SOURCE: Assoc. Rech. Dev. Inf. Chim., Paris, 75005, Fr.
SOURCE: Journal of Chemical Information and Computer Sciences (1985), 25(3), 326-33
CODEN: JCISD8; ISSN: 0095-2338
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The DARC system for documentation and artificial intelligence involving **chem. structural** information is described and its topol. concepts are discussed with respect to interactive data processing systems. Operational realigations of the DARC system are described, including the knowledge **database**, functions of **inference** engines, and interface with users. Computer-aided design applications to the database are detailed in synthesis design, and structure elucidation.

L100 ANSWER 10 OF 45 PASCAL COPYRIGHT 2003 INIST-CNRS. ALL RIGHTS RESERVED. on STN DUPLICATE 1
ACCESSION NUMBER: 2002-0476596 PASCAL
COPYRIGHT NOTICE: Copyright .COPYRGT. 2002 INIST-CNRS. All rights reserved.
TITLE (IN ENGLISH): Molecular **descriptors** that influence the amount of drugs transfer into human breast milk
AUTHOR: AGATONOVIC-KUSTRIN S.; LING L. H.; THAM S. Y.; ALANY R. G.
CORPORATE SOURCE: School of Pharmaceutical, Molecular and Biomedical Science, University of South Australia, North Terrace, Adelaide 5000, Australia; School of Pharmaceutical Sciences, Universiti Sains, Penang 11800, Malaysia; Division of Pharmacy, The University of Auckland, Auckland, New Zealand
SOURCE: Journal of pharmaceutical and biomedical analysis,

(2002), 29(1-2), 103-119, 163 refs.
ISSN: 0731-7085 CODEN: JPBADA
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: Netherlands
LANGUAGE: English
AVAILABILITY: INIST-19962, 354000101592010120

AB Most drugs are excreted into breast milk to some extent and are bioavailable to the infant. The ability to predict the approximate amount of drug that might be present in milk from the drug structure would be very useful in the clinical setting. The aim of this research was to simplify and upgrade the previously developed model for prediction of the milk to plasma (M/P) concentration ratio, given only the molecular structure of the drug. The set of 123 drug compounds, with experimentally derived M/P values taken from the literature, was used to develop, test and validate a predictive model. Each compound was encoded with 71 calculated molecular structure descriptors, including constitutional descriptors, topological descriptors, **molecular** connectivity, geometrical descriptors, quantum **chemical** descriptors, **physicochemical** descriptors and liquid properties. Genetic **algorithm** was used to select a subset of the descriptors that best describe the drug transfer into breast milk and artificial neural network (ANN) to correlate selected descriptors with the M/P ratio and develop a QSAR. The averaged literature M/P values were used as the ANN's output and calculated molecular descriptors as the inputs. A nine-descriptor nonlinear computational neural network model has been developed for the estimation of M/P ratio values for a data set of 123 drugs. The model included the percent of oxygen, parachor, density, highest occupied molecular orbital energy (HOMO), topological indices (.sub.XV2, .sub.X2 and .sub.X1) and shape indices (K3, .kappa.2), as the inputs had four hidden neurons and one output neuron. The QSPR that was developed indicates that molecular size (parachor, density) shape (topological shape indices, molecular connectivity indices) and electronic properties (HOMO) are the most important for drug transfer into breast milk. Unlike previously reported models, the QSPR model described here does not require experimentally derived parameters and could potentially provide a useful prediction of M/P ratio of new drugs only from a sketch of their structure and this approach might also be useful for drug information service. Regardless of the model or method used to estimate drug transfer into breast milk, these predictions should only be used to assist in the evaluation of risk, in conjunction with assessment of the infant's response.

L100 ANSWER 11 OF 45 PASCAL COPYRIGHT 2003 INIST-CNRS. ALL RIGHTS RESERVED.
on STN DUPLICATE 2
ACCESSION NUMBER: 2000-0175430 PASCAL
COPYRIGHT NOTICE: Copyright .COPYRGT. 2000 INIST-CNRS. All rights reserved.
TITLE (IN ENGLISH): Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Development of a **relational database**
AUTHOR: GERNER I.; GRAETSCHEL G.; KAHL J.; SCHLEDE E.
CORPORATE SOURCE: Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV), Thielallee 88-92, 14195 Berlin, Germany, Federal Republic of
SOURCE: ATLA. Alternatives to laboratory animals, (2000), 28(1), 11-28, 14 refs.
ISSN: 0261-1929
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: United Kingdom
LANGUAGE: English

Searched by Barb O'Bryen, STIC 308-4291

AVAILABILITY: INIST-22450, 354000086258330020
AB For new **chemical** substances that are notified within the European Union, data sets have to be submitted to the National Competent Authorities. The data submitted have to demonstrate the **physicochemical** and toxic properties of the new **chemical** , such as solubility, partition coefficients and spectra, as well as acute toxic properties and the potential to cause local irritant or corrosive effects In order to minimise testing for notification purposes (for example, animal testing), it is necessary to develop stepwise assessment procedures, including **structure-activity** considerations, alternative methods (for example, in vitro tests), and computerised **structure-activity** relationship (SAR) models. An electronic **database** was developed which contains **physicochemical** and toxicological data on approximately 1300 **chemical** substances. It is used for regulatory structure-property relationship (SPR) and SAR considerations, and for the development of rules for a decision support system (DSS) for the introduction of alternative methods into local irritancy/corrosivity testing strategies. The information stored in the **database** is derived from proprietary data, so it is not possible to publish the data directly. Therefore, the **database** is evaluated by regulators, and the information derived from the data is used for the development of scientific information about SARs. This information can be published, for example, by means of tables correlating measured **physicochemical** values and specific toxic effects caused by the measured **chemical** , This information is introduced to the public by means of a DSS that predicts local irritant/corrosive potential of a **chemical** by listing so-called exception rules of the kind IF (**physicochemical** property) A THEN not (toxic) Effect B and so-called structural rules of the kind IF Substructure A THEN Effect B. These DSS rules "translate" proprietary data into scientific knowledge that can be published.

L100 ANSWER 12 OF 45 PASCAL COPYRIGHT 2003 INIST-CNRS. ALL RIGHTS RESERVED.
on STN DUPLICATE 3
ACCESSION NUMBER: 1998-0515187 PASCAL
COPYRIGHT NOTICE: Copyright .COPYRGT. 1998 INIST-CNRS. All rights reserved.
TITLE (IN ENGLISH): **Molecular** inference via unidirectional **chemical** reactions
Evolvable systems : from biology to hardware : Lausanne, 23-25 September 1998
AUTHOR: MULAWKA J. J.; OCWIEJA M. J.
SIPPER Moshe (ed.); MANGE Daniel (ed.); PEREZ-URIBE Andres (ed.)
CORPORATE SOURCE: Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland
SOURCE: Lecture notes in computer science, (1998), 1478, 372-379, 16 refs.
Conference: 2 ICES : international conference on evolvable systems, Lausanne (Switzerland), 23 Sep 1998
ISSN: 0302-9743
ISBN: 3-540-64954-9
DOCUMENT TYPE: Journal; Conference
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: Germany, Federal Republic of; United States
LANGUAGE: English
AVAILABILITY: INIST-16343, 354000070103180380
AB Inference process plays an important role in the realisation of expert systems. In this paper it is shown that chemical reactions may by used to perform molecular **inference** according to the **algorithm** of forward chaining. This method is accomplished by an adequate interpretation of inorganic chemical compounds and unidirectional reactions. In our approach premise clauses are represented by the

reactants while conclusion clauses are represented by the products of reaction. Different inorganic compounds and reactions have been discussed with respect to their utility for the molecular inference. Special attention is focused on qualitative chemistry and a number of reactions has been taken into account. Experimental results demonstrating application of these reactions in expert systems are provided.

L100 ANSWER 13 OF 45 PASCAL COPYRIGHT 2003 INIST-CNRS. ALL RIGHTS RESERVED.
on STN DUPLICATE 4

ACCESSION NUMBER: 1997-0268111 PASCAL
COPYRIGHT NOTICE: Copyright .COPYRGT. 1997 INIST-CNRS. All rights reserved.
TITLE (IN ENGLISH): Selecting optimally diverse compounds from structure **databases**: A validation study of two-dimensional and three-dimensional molecular **descriptors**
AUTHOR: MATTER H.
CORPORATE SOURCE: TRIPOS GmbH, Martin-Kollar-Str. 15, 81829 Muenchen, Germany, Federal Republic of
SOURCE: Journal of medicinal chemistry, (1997), 40(8), 1219-1229, 51 refs.
ISSN: 0022-2623 CODEN: JMCMAR
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: United States
LANGUAGE: English
AVAILABILITY: INIST-9165, 354000064961200060

AB The efficiency of the drug discovery process can be significantly improved using design techniques to maximize the diversity of structure **databases** or combinatorial libraries. Here, several **physicochemical** descriptors were investigated to quantify molecular diversity. Based on the 2D or 3D topological similarity of molecules, the relationship between **physicochemical** metrics and **biological** activity was studied to find valid descriptors. Several compounds were selected using those descriptors from a **database** containing diverse templates and 55 **biological** classes. It was evaluated whether the obtained subsets represent all **biological** properties and **structural** variations of the original **database**. In addition, hierarchical cluster analyses were used to group molecules from the parent **database**, which should have similar **biological** properties. Using various sets of **structurally** similar molecules, it was possible to derive quantitative measures for compound similarities in relation to **biological** properties. A similarity radius for 2D fingerprints and molecular steric fields was estimated; compounds within this radius of another **molecule** were shown to have comparable **biological** properties. This study demonstrates that 2D fingerprints alone or in combination with other metrics as the primary descriptor allow to handle global diversity. In addition, standard atom-pair descriptors or molecular steric fields can be used to correlate **structural** diversity with **biological activity** . Hence, the latter two descriptors can be classified as secondary descriptors useful for analog library design, while 2D fingerprints are applicable to design a general library for lead discovery. Based on these findings, an optimally diverse subset containing only 38% of the entire IC93 **database** was generated using 2D fingerprints. Here no structure is more similar than 0.85 to any other (Tanimoto coefficient), but all **biological** classes were selected. This reduction of redundancy led to a child **database** with the same **physicochemical** diversity space, which contains the same information as the original **database**.

L100 ANSWER 14 OF 45 PASCAL COPYRIGHT 2003 INIST-CNRS. ALL RIGHTS RESERVED.

on STN DUPLICATE 5
ACCESSION NUMBER: 1993-0676423 PASCAL
TITLE (IN ENGLISH): **Algorithm** and computer program **Pro** -Anal for analysis of relationship between **structure** and **activity** in a family of proteins or peptidese
AUTHOR: EROSHKIN A. M.; ZHILKIN P. A.; FOMIN V. I.
CORPORATE SOURCE: NPO Vector', res. inst. molecular biology, Novosibirsk 633159, Russian Federation
SOURCE: Computer applications in the biosciences, (1993), 9(5), 491-497, 17 refs.
ISSN: 0266-7061 CODEN: COABER
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: United Kingdom
LANGUAGE: English
AVAILABILITY: INIST-21331, 354000048283990010
AB In this paper we introduce a computer **algorithm** and program Pro-Anal for analysis of the **structure-activity** relationship in a family of evolutionarily related (and/or artificially mutated) proteins/peptides. The program uses aligned amino acid sequences with data of their activity (pK, K.sub.m, ED.sub.5.sub.0 or any other) and searches for correlations between data on activity and various **physico-chemical** characteristics of different regions in primary structures. In automatic mode, the program generates and verifies hypotheses on the disposition of a sequential modulating region in a protein, and key characteristics of the region. In manual mode, users can generate and analyze their own hypotheses. The program is implemented on IBM PC or compatible computers. It is designed to be easily handled by the occasional computer user and yet it is powerful enough for experienced professionals

L100 ANSWER 15 OF 45 PASCAL COPYRIGHT 2003 INIST-CNRS. ALL RIGHTS RESERVED. on STN
ACCESSION NUMBER: 2001-0222782 PASCAL
COPYRIGHT NOTICE: Copyright .COPYRGT. 2001 INIST-CNRS. All rights reserved.
TITLE (IN ENGLISH): Theoretically-derived molecular **descriptors** important in human intestinal absorption
AUTHOR: AGATONOVIC-KUSTRIN S.; BERESFORD R.; YUSOF A. Pauzi M.
CORPORATE SOURCE: School of Pharmaceutical Sciences, Universiti Sains Malaysia, Penang, 11800, Malaysia; School of Pharmacy, University of Otago, P.O. Box 913, Dunedin, New Zealand
SOURCE: Journal of pharmaceutical and biomedical analysis, (2001), 25(2), 227-237, 26 refs.
ISSN: 0731-7085 CODEN: JPBADA
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: Netherlands
LANGUAGE: English
AVAILABILITY: INIST-19962, 354000095065860080
AB A quantitative structure human intestinal absorption relationship was developed using artificial neural network (ANN) modeling. A set of 86 drug compounds and their experimentally-derived intestinal absorption values used in this study was gathered from the literature and a total of 57 global **molecular** descriptors, including constitutional, topological, **chemical**, geometrical and quantum **chemical** descriptors, calculated for each compound. A supervised network with radial basis transfer **function** was used to correlate calculated **molecular** descriptors with experimentally-derived measures of human intestinal absorption. A genetic **algorithm** was then used to select important molecular descriptors. Intestinal absorption values

(1A%) were used as the ANN's output and calculated molecular descriptors as the inputs. The best genetic neural network (GNN) model with 15 input descriptors was chosen, and the significance of the selected descriptors for intestinal absorption examined. Results obtained with the model that was developed indicate that lipophilicity, conformational stability and inter-molecular interactions (polarity, and hydrogen bonding) have the largest impact on intestinal absorption.

L100 ANSWER 16 OF 45 PASCAL COPYRIGHT 2003 INIST-CNRS. ALL RIGHTS RESERVED. on STN

ACCESSION NUMBER: 2002-0155695 PASCAL
TITLE (IN ENGLISH): Issues in predicting **protein** function from sequence
Special issue: Gene Function Part 1
AUTHOR: PONTING Chris P.
DAVIDSON Duncan (introd.); BISHOP Martin (introd.)
CORPORATE SOURCE: MRC Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, South Parks Road, Oxford, OX1 3QX, United Kingdom
MRC Human Genetics Unit, Edinburgh, United Kingdom; UK MRC HGMP Resource Centre, Hinxton, United Kingdom
SOURCE: Briefings in bioinformatics, (2001), 2(1), 19-29, 61 refs.
ISSN: 1467-5463
DOCUMENT TYPE: Journal
BIBLIOGRAPHIC LEVEL: Analytic
COUNTRY: United Kingdom
LANGUAGE: English
AVAILABILITY: INIST-27143

AB Identifying homologues, defined as genes that arose from a common evolutionary ancestor, is often a relatively straightforward task, thanks to recent advances made in estimating the statistical significance of sequence similarities found from **database** searches. The extent by which homologues possess similarities in function, however, is less amenable to statistical analysis. Consequently, predicting function by homology is a qualitative, rather than quantitative, process and requires particular care to be taken. This review focuses on the various approaches that have been developed to predict function from the scale of the atom to that of the organism. Similarities in homologues' functions differ considerably at each of these different scales and also vary for different domain families. It is argued that due attention should be paid to all available clues to function, including orthologue identification, conservation of particular residue types, and the **co-occurrence** of domains in proteins. Pitfalls in **database** searching methods arising from amino acid compositional bias and **database** size effects are also discussed.

L100 ANSWER 17 OF 45 INSPEC (C) 2003 IEE on STN

ACCESSION NUMBER: 1999:6161375 INSPEC
DOCUMENT NUMBER: C1999-03-1230R-031
TITLE: **Molecular** inference via unidirectional **chemical** reactions.
AUTHOR: Mulawka, J.J.; Ocwieja, M.J. (Warsaw Univ. of Technol., Poland)
SOURCE: Evolvable Systems: From Biology to Hardware. Second International Conference, ICES 98 Proceedings
Editor(s): Sipper, M.; Mange, D.; Perez-Uribe, A.
Berlin, Germany: Springer-Verlag, 1998. p.372-9 of ix+382 pp. 16 refs.
Conference: Lausanne, Switzerland, 23-25 Sept 1998
ISBN: 3-540-64954-9
DOCUMENT TYPE: Conference Article
TREATMENT CODE: Practical; Theoretical

Searched by Barb O'Bryen, STIC 308-4291

COUNTRY: Germany, Federal Republic of
LANGUAGE: English
AB Inference process plays an important role in the realisation of expert systems. In this paper it is shown that chemical reactions may by used to perform molecular **inference** according to the **algorithm** of forward chaining. This method is accomplished by an adequate interpretation of inorganic chemical compounds and unidirectional reactions. In our approach premise clauses are represented by the reactants while conclusion clauses are represented by the products of reaction. Different inorganic compounds and reactions have been discussed with respect to their utility for the molecular inference. Special attention is focused on qualitative chemistry and a number of reactions has been taken into account. Experimental results demonstrating application of these reactions in expert systems are provided.

L100 ANSWER 18 OF 45 INSPEC (C) 2003 IEE on STN
ACCESSION NUMBER: 1998:6018364 INSPEC
DOCUMENT NUMBER: A9820-3115-006; C9810-7320-075
TITLE: Construction of a combined version of a JSM-type plausible **reasoning** system reflecting the quantum-**chemical** properties of **molecules.**
AUTHOR: D'yachkov, P.N.; Manevich, S.I.; Putrin, A.V.; Finn, V.K.
SOURCE: Automatic Documentation and Mathematical Linguistics (1997) vol.31, no.2, p.16-23. 13 refs.
Published by: Allerton Press
Price: CCCC 0005-1055/97/$50.00
CODEN: ADMLAE ISSN: 0005-1055
SICI (Trl): 0005-1055(1997)31:2L.16:CCVT;1-J
Translation of: Nauchno-Tekhnicheskaya Informatsiya, Seriya 2 (1997) no.3, p.16-23. 13 refs.
CODEN: NIPSBP ISSN: 0548-0027
SICI: 0548-0027(1997)3L.16;1-T
DOCUMENT TYPE: Journal; Translation Abstracted
TREATMENT CODE: Theoretical
COUNTRY: Russian Federation; United States
LANGUAGE: English
AB The capabilities and advantages of an integrated system for predicting the properties of chemical compounds constructed in simultaneous application of the structural and **physicochemical** properties of compounds are analyzed. The construction of such a system for prediction of counter-productive properties of chemical compounds is considered on the basis of a single numerical characteristic-activation energy-taking into account the structural formulas of molecules and of derivatives of molecules. One possible mathematical model for combined application of different parameters within the framework of a plausible **inference** system of certain properties under investigation is described; a procedure is presented for application of such a combined model in a JSM-system for automatic generation of hypotheses and questions, whether the questions have already been solved through use of the JSM-system for the particular model as well as questions that remain to be solved.

L100 ANSWER 19 OF 45 INSPEC (C) 2003 IEE on STN
ACCESSION NUMBER: 1991:3768244 INSPEC
DOCUMENT NUMBER: A91002683; C91005668
TITLE: Elaboration of computer data **bank** for **physicochemical** gas dynamics.
AUTHOR: Losev, S.A.; Shatalov, O.P. (Inst. of Mech., Lomonosov State Univ., Moscow, USSR)
SOURCE: Soviet Journal of Chemical Physics (1990) vol.6, no.12, p.3299-335. 73 refs.
CODEN: SJCPDF ISSN: 0733-2831

Searched by Barb O'Bryen, STIC 308-4291

Translation of: Khimicheskaya Fizika. 73 refs.
CODEN: KHFID9 ISSN: 0207-401X
DOCUMENT TYPE: Journal; Translation Abstracted
TREATMENT CODE: Bibliography; General Review
COUNTRY: USSR; United Kingdom
LANGUAGE: English
AB The paper describes the purpose and structure of an automated dataware system for gas dynamics with recommendations including reliability estimates (ADGDRE). The system consists of a data bank, a generator of simulation of the medium, a library of program modules, and a constructor of program modules. The **physicochemical** data bank consists of four bases, namely initial information, data preparation, recommended data, and model bases. The content of the base of recommended data is described using an example of a base of data on **chemical** reaction rate constants for **molecules** consisting of nitrogen and oxygen atoms. This **database** includes all the reactions between these molecules that are mentioned in the literature. A detailed discussion is devoted to the recommended data on dissociation and recombination reactions of diatomic molecules N2, O2.

L100 ANSWER 20 OF 45 INSPEC (C) 2003 IEE on STN
ACCESSION NUMBER: 1989:3263662 INSPEC
DOCUMENT NUMBER: C89004432
TITLE: Statistical analysis of quantitative **structure activity** relationships (QSAR) in toxicology based on a **relational** data model.
AUTHOR: Weber, E.; Kinscherf, S. (Deutsches Krebsforschungszentrum, Heidelberg, West Germany); von der Trenck, K.T.
SOURCE: Statistical Software Newsletter (Aug. 1988) vol.14, no.2, p.82-8. 12 refs.
CODEN: SSNEEX ISSN: 0173-5896
DOCUMENT TYPE: Journal
TREATMENT CODE: Application
COUNTRY: Germany, Federal Republic of
LANGUAGE: English
AB The analysis of QSARs in toxicology makes use of structural features and **physicochemical** parameters and is aimed at several marks such as the prediction of toxicity, the preliminary assessment of risk, or the validation of alternatives to animal experiments, etc. These various tasks make different requirements for the statistical models for data analysis as well as for the techniques to extract problem-specific data from the **databases**. Consequently, the evaluative routines should be adapted to the **database**. The **structure** and the contents of the **biological database** are outlined and the lateral communication with a spectral **database** is indicated. A flexible management of the **database** and the extraction of information from it require further utilities that are afforded by APL2-mediated extensions of the system. The package TRAINS permits the user-friendly and time-saving application of the complicated structure. Based on these features, the essentials of a flexible system for the data evaluation and its realization are described. The particulars of the arising numerical problems and their solution with the aid of APL2 are extensively treated. The article concludes with the enumeration of further objectives to be achieved.

L100 ANSWER 21 OF 45 LIFESCI COPYRIGHT 2003 CSA on STN
ACCESSION NUMBER: 2002:101298 LIFESCI
TITLE: Predicting **Protein** Cellular Localization Using a Domain Projection Method
AUTHOR: Mott, R.; Schultz, J.; Bork, P.; Ponting, C.P.
CORPORATE SOURCE: Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, United Kingdom; E-mail: rmott@well.ox.ac.uk

SOURCE: Genome Research [Genome Res.], (20020800) vol. 12, no. 8, pp. 1168-1174.
ISSN: 1054-9803.
DOCUMENT TYPE: Journal
FILE SEGMENT: G
LANGUAGE: English
SUMMARY LANGUAGE: English
AB We investigate the **co-occurrence** of domain families in eukaryotic proteins to predict protein cellular localization. Approximately half (300) of SMART domains form a "small-world network", linked by no more than seven degrees of separation. Projection of the domains onto two-dimensional space reveals three clusters that correspond to cellular compartments containing secreted, cytoplasmic, and nuclear proteins. The projection method takes into account the existence of "bridging" domains, that is, instances where two domains might not occur with each other but frequently **co-occur** with a third domain; in such circumstances the domains are neighbors in the projection. While the majority of domains are specific to a compartment ("locale"), and hence may be used to localize any protein that contains such a domain, a small subset of domains either are present in multiple locales or occur in transmembrane proteins. Comparison with previously annotated proteins shows that SMART domain data used with this approach can predict, with 92% accuracy, the localizations of 23% of eukaryotic proteins. The coverage and accuracy will increase with improvements in domain **database** coverage. This method is complementary to approaches that use amino-acid composition or identify sorting sequences; these methods may be combined to further enhance prediction accuracy.

L100 ANSWER 22 OF 45 LIFESCI COPYRIGHT 2003 CSA on STN
ACCESSION NUMBER: 1999:45436 LIFESCI
TITLE: Functional Sites in **Pro**- and Eukaryotic Genomes: Computer Models for Predicting Activity
AUTHOR: Kolchanov, N.A.; Ponomarenko, M.P.; Ponomarenko, Y.V.; Podkolodnyi, N.L.; Frolov, A.S.
CORPORATE SOURCE: Institute of Cytology and Genetics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia; E-mail: kol@bionet.nsc.ru
SOURCE: Molecular Biology [Mol. Biol.], (19980400) vol. 32, no. 2, pp. 255-267.
ISSN: 0026-8933.
DOCUMENT TYPE: Journal
FILE SEGMENT: G
LANGUAGE: English
SUMMARY LANGUAGE: English
AB Here we propose an approach for predicting the activity of functional DNA and RNA sites. This approach includes (1) identification of context-dependent conformational, **physicochemical**, and statistical properties of sites significant for their functioning; (2) development of a model on their basis for predicting site activity from its sequence; and (3) automatic generation of programs for predicting site activity based on these models. This approach has been realized as a computer system ACTIVITY, which includes **databases** of site activity as well as conformational, **physicochemical**, and statistical properties of DNA and RNA. ACTIVITY is accessible via Internet (http://www.bionet.nsc.ru/SRCG/Activity/) and allows real-time analysis of experimental data on functional site activity. We analyzed 70 samples of sites involved in various **molecular biological processes** and revealed statistical, conformational, and **physicochemical** properties significant for activity of these sites. We also developed methods for predicting site activity from their nucleotide sequences.

L100 ANSWER 23 OF 45 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN

ACCESSION NUMBER: 1999:263629 BIOSIS
DOCUMENT NUMBER: PREV199900263629
TITLE: From **fold** predictions to function predictions: Automation of functional site conservation analysis for functional genome predictions.
AUTHOR(S): Zhang, Baohong; Rychlewski, Leszek; Pawlowski, Krzysztof; Fetrow, Jacquelyn S.; Skolnick, Jeffrey; Godzik, Adam [Reprint author]
CORPORATE SOURCE: The Burnham Institute, 10901 North Torrey Pines Rd., La Jolla, CA, 92037, USA
SOURCE: Protein Science, (May, 1999) Vol. 8, No. 5, pp. 1104-1115. print.
ISSN: 0961-8368.
DOCUMENT TYPE: Article
LANGUAGE: English
ENTRY DATE: Entered STN: 15 Jul 1999
Last Updated on STN: 15 Jul 1999

AB A **database** of **functional** sites for proteins with known **structures**, SITE, is constructed and used in conjunction with a simple pattern matching program SiteMatch to evaluate possible function conservation in a recently constructed **database** of fold predictions for Escherichia coli proteins (Rychlewski L et al., 1999, Protein Sci 8:614-624). In this and other prediction **databases**, fold predictions are based on **algorithms** that can recognize weak sequence similarities and putatively assign new proteins into already characterized protein families. It is not clear whether such sequence similarities arise from distant homologies or general similarity of **physicochemical** features along the sequence. Leaving aside the important question of nature of relations within fold superfamilies, it is possible to assess possible function conservation by looking at the pattern of conservation of crucial functional residues. SITE consists of a multilevel **function** description based on **structure** annotations and structure analyses. In particular, active site residues, ligand binding residues, and patterns of hydrophobic residues on the protein surface are used to describe different functional features. SiteMatch, a simple pattern matching program, is designed to check the conservation of residues involved in protein activity in alignments generated by any alignment method. Here, this procedure is used to study conservation of functional features in alignments between protein sequences from the E. coli genome and their optimal structural templates. The optimal templates were identified and alignments taken from the **database** of genomic structural predictions was described in a previous publication (Rychlewski L et al., 1999, Protein Sci 8:614-624). An automated assessment of function conservation is used to analyze the relation between fold and function similarity for a large number of fold predictions. For instance, it is shown that identifying low significance predictions with a high level of functional residue conservations can be used to extent the prediction sensitivity for fold prediction methods. Over 100 new fold/function predictions in this class were obtained in the E. coli genome. At the same time, about 30% of our previous fold predictions are not confirmed as function predictions, further highlighting the problem of function divergence in fold superfamilies.

L100 ANSWER 24 OF 45 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 1997:249060 BIOSIS
DOCUMENT NUMBER: PREV199799548263
TITLE: A **bank** of protein family patterns for rapid identification of possible functions of amino acid sequences.
AUTHOR(S): Bachinsky, A. G. [Reprint author]; Yarigin, A. A.; Guseva, E. H.; Kulichkov, V. A.; Nizolenko, L. P.
CORPORATE SOURCE: Theoretical Dep., Research Inst. Molecular Biol., SRC VB 'Vector', Koltsovo, Novosibirsk Region 633159, Russia

SOURCE: Computer Applications in the Biosciences, (1997) Vol. 13, No. 2, pp. 115-122.
CODEN: COABER. ISSN: 0266-7061.
DOCUMENT TYPE: Article
(Software)
LANGUAGE: English
ENTRY DATE: Entered STN: 13 Jun 1997
Last Updated on STN: 13 Jun 1997

AB A method and software tool to develop patterns of protein families has been designed. These patterns are intended for the identification of local similarities in arbitrary amino acid sequences with proteins of the SWISS-PROT bank. The method is based on the physical, **chemical** and **structural** properties of amino acids. It assembles a 'best set' of elements (a pattern) for a given group of aligned related proteins. These elements provide discrimination between proteins of a family and representatives of other families or random sequences. The method combines the advantages of BLOCKS (automatic generation of multiple elements for protein groups), PROSITE (simplicity of element presentation) and matrices/profiles (different distinctions between amino acids for different positions of aligned sequences). Using our method, a data bank of protein family patterns, PROF-PAT, is produced. This data bank is based on the 27 752 amino acid sequences of SWISS-PROT bank release 24. The characteristics of patterns of 743 related protein groups are described. The results of comparisons of PROF-PAT patterns with the proteins of the SWISS-PROT bank are discussed.

L100 ANSWER 25 OF 45 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 1992:284928 BIOSIS
DOCUMENT NUMBER: PREV199294009578; BA94:9578
TITLE: **NON-SEQUENCE DATABASES** FOR **BIOLOGICAL** ACTIVITY AND **PHYSICOCHEMICAL** PROPERTIES.
AUTHOR(S): JONES C S [Reprint author]; TSUGITA A; SATAKE K; OKIBAYASHI F; IMAI K; YAGI T; TAKAHASHI K; YEH L-S
CORPORATE SOURCE: JPN INT PROTEIN INFORMATION DATABASE RES INST BIOSCI, SCI UNIV TOKYO, YAMAZAKI, NODA 278, JPN
SOURCE: Protein Sequences and Data Analysis, (1991) Vol. 4, No. 6, pp. 367-374.
CODEN: PSDAE6. ISSN: 0931-9506.
DOCUMENT TYPE: Article
FILE SEGMENT: BA
LANGUAGE: ENGLISH
ENTRY DATE: Entered STN: 10 Jun 1992
Last Updated on STN: 10 Jun 1992

AB A **biological** activity **database** and a **physicochemical** property **database** are described. They are intended to complement the protein sequence **database** of PIR-International. The **Biological** Activity **Database** and the **Physicochemical** Property **Database** contain information regarding the **biological** activity and the **physicochemical** properties of proteins, respectively. In addition they also provide information about wild-type molecules with which information concerning variant molecules may be compared. Data on artificial variant molecules are stored in the Artificial Variant **Database** which is described separately.

L100 ANSWER 26 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2003:728055 SCISEARCH
THE GENUINE ARTICLE: 710XG
TITLE: Neighborhood behavior. Fuzzy molecular **descriptors** and their influence on the relationship between structural similarity and property similarity
AUTHOR: Horvath D (Reprint); Mao B

CORPORATE SOURCE: CEREP SA, 128, Rue Danton, F-92506 Rueil Malmaison, France (Reprint); CEREP SA, F-92506 Rueil Malmaison, France; CEREP Inc, Redmond, WA USA
COUNTRY OF AUTHOR: France; USA
SOURCE: QSAR & COMBINATORIAL SCIENCE, (JUL 2003) Vol. 22, No. 5, pp. 498-509.
Publisher: WILEY-V C H VERLAG GMBH, PO BOX 10 11 61, D-69451 WEINHEIM, GERMANY.
ISSN: 1611-020X.
DOCUMENT TYPE: Article; Journal
LANGUAGE: English
REFERENCE COUNT: 9

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The similarity principle, stating that molecules of similar structure behave similarly, is an important concept in medicinal **chemistry**. A properly characterized and well understood neighborhood behavior of the **structural** space versus the **activity** space is fundamental for the application of the similarity principle in computational **chemistry**. In this work we focus on the utilization of a fuzzy pharmacophore description of molecular similarity and specifically on the influence of fuzzy pharmacophore pattern matching on the neighborhood behavior (NB) of the similarity scoring scheme. NB is defined as a **structure activity** relationship between the intermolecular distances/ dissimilarities in the pharmacophore fingerprint **structure** space and the corresponding **activity** differences, formally seen as intermolecular distances in the activity spaces. The latter are defined on hand of a wide variety of datasets on pharmacological and **physico-chemical** properties and property profiles. We also investigate the clustering behavior (CB), where the **structure-activity** relationship is described in terms of distance-derived associations of compounds into clusters via classical hierarchical clustering procedures. The neighborhood behavior and the cluster behavior provide alternative and complementary criteria for evaluating the pertinence of a molecular similarity metric.

L100 ANSWER 27 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2003:265915 SCISEARCH
THE GENUINE ARTICLE: 655FP
TITLE: MATRIX, a new **algorithm** for predicting **biological activity** of organic **molecules** based on multidimensional analysis of **physicochemical descriptors** of modern pharmaceuticals: I. General principles
AUTHOR: Pogrebnyak A V (Reprint); Oganesyan E T; Glushko A A
CORPORATE SOURCE: Pyatigorsk State Pharmaceut Acad, Pyatigorsk 357500, Russia (Reprint)
COUNTRY OF AUTHOR: Russia
SOURCE: RUSSIAN JOURNAL OF ORGANIC CHEMISTRY, (NOV 2002) Vol. 38, No. 11, pp. 1564-1575.
Publisher: MAIK NAUKA/INTERPERIODICA, C/O KLUWER ACADEMIC-PLENUM PUBLISHERS, 233 SPRING ST, NEW YORK, NY 10013-1578 USA.
ISSN: 1070-4280.
DOCUMENT TYPE: Article; Journal
LANGUAGE: English
REFERENCE COUNT: 32

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The quantum-**chemical** calculation of **structures** of organic **molecules** belonging to 1067 modern pharmaceuticals was carried out by semiempirical (AM1, PM3, MNDO, CNDO/2, MINDO/3) and ab initio (6-31G) procedures taking into account the hydration effects. Each molecule was characterized by 149 topochemical and quantum-

chemical descriptors. Basing on combination of multidimensional analysis procedures a new method was developed for forecasting the **biological** activity of organic compounds consisting in determination of proximity of the molecules on a surface of a potential function in the multidimensional space of descriptors (MATRIX **algorithm**).

L100 ANSWER 28 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2002:574200 SCISEARCH
THE GENUINE ARTICLE: 570ED
TITLE: Correlation properties of the autocorrelation **descriptor** for molecules
AUTHOR: Hollas B (Reprint)
CORPORATE SOURCE: Univ Ulm, Dept Theoret Comp Sci, D-89069 Ulm, Germany (Reprint)
COUNTRY OF AUTHOR: Germany
SOURCE: MATCH-COMMUNICATIONS IN MATHEMATICAL AND IN COMPUTER CHEMISTRY, (MAR 2002) No. 45, pp. 27-33.
Publisher: UNIV BAYREUTH, DEPT MATHEMATICS, C/O PROF DR A KERBER, D-95440 BAYREUTH, GERMANY.
ISSN: 0340-6253.
DOCUMENT TYPE: Article; Journal
LANGUAGE: English
REFERENCE COUNT: 15

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The autocorrelation descriptor is a molecular descriptor encoding both **molecular structure** and **physico-chemical** properties attributed to atoms as a vector. Applications include QSAR studies and screening of large **databases**. Using random graphs, we show that the autocorrelation descriptor may contain highly redundant information even if the encoded properties are independent. We show that this shortcoming can easily be eliminated by centering properties, facilitating subsequent statistical analysis of the generated data.

L100 ANSWER 29 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2000:377956 SCISEARCH
THE GENUINE ARTICLE: 313KF
TITLE: Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular **descriptors**
AUTHOR: Gramatica P (Reprint); Corradi M; Consonni V
CORPORATE SOURCE: UNIV INSUBRIA, DEPT STRUCT & FUNCT BIOL, QSAR RES UNIT, VIA DUNANT 3, I-21100 VARESE, ITALY (Reprint); UNIV MILANO BICOCCA, DEPT ENVIRONM SCI, MILANO CHEMOMETR & QSAR RES GRP, I-20126 MILAN, ITALY
COUNTRY OF AUTHOR: ITALY
SOURCE: CHEMOSPHERE, (SEP 2000) Vol. 41, No. 5, pp. 763-777.
Publisher: PERGAMON-ELSEVIER SCIENCE LTD, THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, ENGLAND.
ISSN: 0045-6535.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: AGRI
LANGUAGE: English
REFERENCE COUNT: 60

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Soil sorption coefficients (K-OC) of 185 non-ionic organic heterogeneous pesticides have been studied searching for quantitative **structure**-property relationships (QSPRs). The **chemical** description of pesticide **structure** has been made in terms of some molecular descriptors: count descriptors, topological indices, information indices, fragment-based descriptors and weighted holistic invariant molecular (WHIM) descriptors; these last are statistical indices

describing size, shape, symmetry and atom distribution of molecules in the three-dimensional space. Three new topological indices derived from the electrotopological state indices of Kier and Hall were proposed. Multiple linear regression analysis was performed after previous selection of the descriptors mostly correlated to the response by Genetic **Algorithms**. The obtained results confirm the capability of the proposed approach to give predictive models for one of the most important partition properties, such as soil sorption coefficient (K-OC). (C) 2000 Elsevier Science Ltd. All rights reserved.

L100 ANSWER 30 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2000:406799 SCISEARCH
THE GENUINE ARTICLE: 317GJ
TITLE: Binary formal **inference**-based recursive modeling using multiple atom and **physicochemical** property class pair and torsion descriptors as decision criteria
AUTHOR: Cho S J (Reprint); Shen C F; Hermsmeier M A
CORPORATE SOURCE: BRISTOL MYERS SQUIBB CO, COMBINATORIAL DRUG DISCOVERY, 5 RES PKWY, WALLINGFORD, CT 06492 (Reprint); BRISTOL MYERS SQUIBB CO, NONCLIN BIOSTAT, PRINCETON, NJ 08543; BRISTOL MYERS SQUIBB CO, COMBINATORIAL DRUG DISCOVERY, PRINCETON, NJ 08543
COUNTRY OF AUTHOR: USA
SOURCE: JOURNAL OF CHEMICAL INFORMATION AND COMPUTER SCIENCES, (MAY-JUN 2000) Vol. 40, No. 3, pp. 668-680.
Publisher: AMER CHEMICAL SOC, 1155 16TH ST, NW, WASHINGTON, DC 20036.
ISSN: 0095-2338.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: PHYS
LANGUAGE: English
REFERENCE COUNT: 63
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Analysis of a large amount of information, typically generated by high-throughput screening, is a very difficult task. To address this problem, we have developed binary formal **inference**-based recursive modeling using atom and **physicochemical** property class pair and torsion descriptors. Recursive partitioning is an exploratory technique for identifying structure in data. The implemented **algorithm** utilizes a statistical hypothesis resting, similar to Hawkins' formal **inference**-based recursive modeling program, to separate a data set into two homogeneous subsets at each splitting **node**. This process is repented recursively until no further separation can occur. Our implementation of recursive partitioning differs from previously reported approaches by employing a method to extract multiple features at each splitting **node**. The method was examined for its ability to distinguish random and real data sets. The effect of including a single descriptor and multiple descriptors in the splitting descriptor set was also studied. The method was tested using 27 401 National Cancer Institute (NCI) compounds and their pGI50 (-log(GI(50))) against the NCl-H23 cell line. The analyses show that partitioning using multiple descriptors is advantageous in analyzing the **structure-activity** relationship information.

L100 ANSWER 31 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2000:596657 SCISEARCH
THE GENUINE ARTICLE: 339WT
TITLE: Prediction of drug transfer into human milk from theoretically derived **descriptors**
AUTHOR: AgatonovicKustrin A (Reprint); Tucker I G; Zecevic M; Zivanovic L J
CORPORATE SOURCE: UNIV OTAGO, SCH PHARM, POB 913, DUNEDIN, NEW ZEALAND (Reprint); UNIV BELGRADE, FAC PHARM, YU-11000 BELGRADE,

SERBIA, YUGOSLAVIA
COUNTRY OF AUTHOR: NEW ZEALAND; YUGOSLAVIA
SOURCE: ANALYTICA CHIMICA ACTA, (9 AUG 2000) Vol. 418, No. 2, pp. 181-195.
Publisher: ELSEVIER SCIENCE BV, PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS.
ISSN: 0003-2670.
DOCUMENT TYPE: General Review; Journal
FILE SEGMENT: PHYS
LANGUAGE: English
REFERENCE COUNT: 110

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The goal of this study was to develop a genetic neural network (GNN) model to predict the degree of drug transfer into breast milk, depending on the molecular structure descriptors, and to compare it with the current model. A supervised network with back-propagation learning rule and multilayer perceptron (MLP) architecture was used to correlate activity with descriptors that were preselected by a genetic **algorithm**. The set of 60 drug compounds and their experimentally derived MIP values used in this study were gathered from Literature. A total of 61 calculated **structural** features including constitutional, topological, **chemical**, geometrical and quantum **chemical** descriptors were generated for each of the 60 compounds. The MIP Values were used as the ANNs output and calculated molecular descriptors as the inputs.

The best GNN model with 26 input descriptors is presented, and the **chemical** significance of the chosen descriptors is discussed. Strong correlation of predicted versus experimentally derived M/P values (R-2>0.96) for the best ANN model (26-5-5-1) confirms that there is a link between structure and MIP values. The strength of the link is measured by the quality of the external prediction set. With the RMS error of 0.425 and a good visual plot, the external prediction set ensures the quality of the model.

Unlike previously reported models, the GNN model described here does not require experimental parameters and could potentially provide useful prediction of M/P ratio of new potential drugs and reduce the need for actual compound synthesis and experimental M/P ratio determination.

L100 ANSWER 32 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2000:217924 SCISEARCH
THE GENUINE ARTICLE: 293DJ
TITLE: **PHYSEAN**: PHYsical SEquence ANalysis for the identification of protein domains on the basis of physical and **chemical** properties of amino acids
AUTHOR: Ladunga I (Reprint)
CORPORATE SOURCE: SMITHKLINE BEECHAM PHARMACEUT, BIOINFORMAT DEPT, KING OF PRUSSIA, PA 19406 (Reprint); HUNGARIAN ACAD SCI, RES GRP EVOLUTIONARY GENET, H-1051 BUDAPEST, HUNGARY; LORAND EOTVOS UNIV, H-1051 BUDAPEST, HUNGARY
COUNTRY OF AUTHOR: USA; HUNGARY
SOURCE: BIOINFORMATICS, (DEC 1999) Vol. 15, No. 12, pp. 1028-1038.
Publisher: OXFORD UNIV PRESS, GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND.
ISSN: 1367-4803.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 70

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Motivation: PHYSEAN predicts protein classes with highly variable sequences on the basis of their physical, **chemical** and **biological** characteristics such as diverse hydrophobicity, **structural** propensity and steric properties. These

characteristics, calculated from multiple positions in a sequence, may be conserved even between sequences that fail to produce alignments at any acceptable level of statistical significance. PHYSEAN complements methods that require sequence alignments (BLAST, FASTA, dynamic programming) by adding less residue- and position-specific **physicochemical** information on the protein or the domain.

Results: We predict proteins or their domains like signal peptides using physical, **chemical**, geometric, and **biological** properties of the 20 amino acids. This comprehensive set of properties may cover the diagnostic **functional** and **structural** aspects of a domain or a protein class. We automatically select and weight a subset of properties so as to discriminate between, e.g., signal peptides and amino-termini of cytosolic proteins with the lowest number of incorrect predictions. This optimal selection of properties and their weights significantly decreases the number of incorrect predictions as compared to any single property or any combination of unweighted properties. Weights have been optimized by high-performance linear programming models that systematically find the optimal solution from among an astronomic number of property/weight combinations. PHYSEAN's performance is demonstrated by highly accurate predictions of signal peptides (the vehicles for protein transport across membranes) and their cleavage sites. The results indicate reliable predictions are possible even in the lack of sequence conservation using an automated physical and **chemical** analysis of proteins.

L100 ANSWER 33 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 1998:842635 SCISEARCH
THE GENUINE ARTICLE: 133PZ
TITLE: A validation study of molecular **descriptors** for the rational design of peptide libraries
AUTHOR: Matter H (Reprint)
CORPORATE SOURCE: HOECHST MARION ROUSSEL, COMPUTAT CHEM, CORE RES FUNCT, BLDG G 838, , D-65926 FRANKFURT, GERMANY (Reprint)
COUNTRY OF AUTHOR: GERMANY
SOURCE: JOURNAL OF PEPTIDE RESEARCH, (OCT 1998) Vol. 52, No. 4, pp. 305-314.
Publisher: MUNKSGAARD INT PUBL LTD, 35 NORRE SOGADE, PO BOX 2148, DK-1016 COPENHAGEN, DENMARK.
ISSN: 1397-002X.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 53

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Important molecular descriptors used for establishing quantitative **structure-activity** relationships are investigated to classify similar versus dissimilar peptides. When searching new lead structures, synthesizing and testing compounds which are too similar wastes time and resources. In contrast, any lead optimization program requires the investigation of similar compounds to that lead. Thus, it is important to maximize or minimize the structural diversity of peptides to design useful compound libraries for lead finding or lead refinement projects.

If a molecular descriptor is a useful measure of similarity for the design of peptide libraries, small differences in this descriptor for a pair of **molecules** should only translate into small **biological** differences. Using this paradigm as a basis for descriptor validation, it was possible to rank different molecular descriptors. Those **physicochemical** descriptors are 2D fingerprints and five experimentally or theoretically derived principal property scales. Some theoretically derived metrics are obtained by computing interaction energies or similarity indices on predefined 3D grid points using canonical conformations for individual amino acids. The

resulting 3D data matrices are analyzed using a principal component analysis leading to three principal properties for CoMFA (Comparative Molecular Field Analysis) or CoMSIA (Comparative Molecular Similarity Index Analysis) derived molecular fields.

The descriptor validation results reveal the applicability of design tools on peptide data sets. Experimentally derived descriptors, in general, are more acceptable than computationally derived metrics, while the latter provide a statistically valid alternative to characterize novel building blocks. The CoMSIA metrics perform slighly better than the CoMFA-based principal properties, while GRID-based descriptors are always less acceptable.

L100 ANSWER 34 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 1998:373608 SCISEARCH
THE GENUINE ARTICLE: ZM695
TITLE: Functional sites in **pro-** and eukaryotic genomes: Computer models for predicting activity
AUTHOR: Kolchanov N A (Reprint); Ponomarenko M P; Ponomarenko Y V; Podkolodnyi N L; Frolov A S
CORPORATE SOURCE: RUSSIAN ACAD SCI, INST CYTOL & GENET, NOVOSIBIRSK 630090, RUSSIA (Reprint); RUSSIAN ACAD SCI, CTR COMP, NOVOSIBIRSK 630098, RUSSIA
COUNTRY OF AUTHOR: RUSSIA
SOURCE: MOLECULAR BIOLOGY, (MAR-APR 1998) Vol. 32, No. 2, pp. 220-232.
Publisher: PLENUM PUBL CORP, CONSULTANTS BUREAU, 233 SPRING ST, NEW YORK, NY 10013.
ISSN: 0026-8933.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 49

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Here we propose an approach for predicting the activity of functional DNA and RNA sites. This approach includes (1) identification of context-dependent conformational, **physicochemical**, and statistical properties of sites significant for their functioning; (2) development of a model on their basis for predicting site activity from its sequence; and (3) automatic generation of programs for predicting site activity based on these models. This approach has been realized as a computer system ACTIVITY, which includes **databases** of site activity as well as conformational, **physicochemical**, and statistical properties of DNA and RNA. ACTIVITY is accessible via Internet (http://www.bionet.nsc.ru/SRCG/Activity/) and allows real-time analysis of experimental data on functional site activity. We analyzed 70 samples of sites involved in various **molecular biological processes** and revealed statistical, conformational, and **physicochemical** properties significant for activity of these sites. We also developed methods for predicting site activity from their nucleotide sequences.

L100 ANSWER 35 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 1998:500072 SCISEARCH
THE GENUINE ARTICLE: ZV969
TITLE: 3D-modelling and prediction by WHIM **descriptors.** Part 9. Chromatographic relative retention time and **physico-chemical** properties of polychlorinated biphenyls (PCBs)
AUTHOR: Gramatica P (Reprint); Navas N; Todeschini R
CORPORATE SOURCE: UNIV MILAN, DEPT ENVIRONM SCI, VIA EMANUELI 15, I-20126 MILAN, ITALY (Reprint)
COUNTRY OF AUTHOR: ITALY
SOURCE: CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS, (MAY 1998

) Vol. 40, No. 1, pp. 53-63.
Publisher: ELSEVIER SCIENCE BV, PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS.
ISSN: 0169-7439.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: PHYS
LANGUAGE: English
REFERENCE COUNT: 39

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB **Physico-chemical** properties of polychlorinated biphenyls (PCBs) congeners have been extensively studied searching for quantitative **structure**-property relationships (QSPR). The **chemical** description of PCBs **structure** is made in terms of WHIM descriptors, which are 3D molecular descriptors taking into account size, shape, symmetry and atom distribution of the molecules. The regression models have been obtained by optimizing their prediction power and by selecting the best subset of descriptors by genetic **algorithm**. The results confirm the capability of this approach to give predictive models for important **physico-chemical** properties, such as relative retention time, log K-ow, melting point, total surface area, Henry's law constant, solubility, and aqueous activity coefficients. (C) 1998 Elsevier Science B.V. All rights reserved.

L100 ANSWER 36 OF 45 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN

ACCESSION NUMBER: 97:365219 SCISEARCH
THE GENUINE ARTICLE: WX492
TITLE: 3D-modelling and prediction by WHIM **descriptors** .6. Application of WHIM **descriptors** in QSAR studies
AUTHOR: Todeschini R (Reprint); Gramatica P
CORPORATE SOURCE: DEPT ENVIRONM SCI, VIA EMANUELI 15, I-20126 MILAN, ITALY (Reprint)
COUNTRY OF AUTHOR: ITALY
SOURCE: QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS, (APR 1997) Vol. 16, No. 2, pp. 120-125.
Publisher: VCH PUBLISHERS INC, 303 NW 12TH AVE, DEERFIELD BEACH, FL 33442-1788.
ISSN: 0931-8771.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE
LANGUAGE: English
REFERENCE COUNT: 18

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Three-dimensional molecular indices (WHIM descriptors), proposed in Part 5 [1] are used to search for quantitative **structure-activity** relationships to investigate the **physico-chemical** properties and **biological** activities of different classes of environmental important compounds.

Chlorobenzenes are studied for their interesting **physico-chemical** properties, e.g., melting and boiling points, solubility, lipophilicity (logK(ow)), bioconcentration factor (BCF), and for toxicity (Microtex test and algae). The antagonism of N,N-dimethyl-2-halophenethylamines to epinephrine and histamine is successfully modelled and compared with other models in the literature. Finally, good QSAR models are obtained for modelling the receptor binding affinities (RE) and inductions of aryl hydrocarbon hydroxylase (AHH) for some dioxin analogue compounds, polyhalogenated aryl derivatives

All the obtained models confirm the high modelling power of the WHIM descriptors.

L100 ANSWER 37 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER: 2003-019137 [01] WPIDS
DOC. NO. NON-CPI: N2003-014659

DOC. NO. CPI: C2003-004849
TITLE: Quantitative **structure** property **activity** relationship (QSPAR) generation method for chemical **structure**/biological **activity** research, involves generating QSPAR model and selecting associative significant **descriptors**.
DERWENT CLASS: **B04 T01**
INVENTOR(S): KERI, G; KOEVESDI, I; OERFI, L
PATENT ASSIGNEE(S): (AXXI-N) AXXIMA PHARM AG
COUNTRY COUNT: 100
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2002082329	A2	20021017	(200301)*	EN	48

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2002082329	A2	WO 2002-EP3622	20020402

PRIORITY APPLN. INFO: US 2001-285222P 20010423; EP 2001-108737 20010406

AB WO 200282329 A UPAB: 20030101
NOVELTY - A **database** containing molecular descriptors especially 2D and 3D biological, chemical or physical data is established.
DETAILED DESCRIPTION - A **database** containing molecular descriptors especially 2D and 3D biological, chemical or physical data is established. A model is provided for generating quantitative **structure** property **activity** relationship (QSPAR) and significant descriptors are selected in accordance to their influence to the QSPAR. The model is verified by using a quality parameter and the process of generation of the relationships is continued until the parameter reaches a predetermined value.
INDEPENDENT CLAIMS are also included for:
(1) QSPAR generation system; and
(2) Computer program product storing QSPAR generation instructions.
USE - For chemical **structure**/biological **activity** research, especially for generating quantitative **structure** property **activity** relationship (QSPAR) between structure of chemical compounds and their pharmacological activity for prophylaxis and for treatment of various diseases.
ADVANTAGE - The validated QSPAR model efficiently provides true relationships between the structure of chemical compounds and their pharmacological activity.
DESCRIPTION OF DRAWING(S) - The figure shows the flow diagram illustrating QSPAR generation method.
Dwg.1B/13

L100 ANSWER 38 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-669064 [63] WPIDS
DOC. NO. CPI: C2003-182231
TITLE: Molecular designing for developing a drug, comprises

analyzing a three dimensional quantitative **structure activity** relation to estimate a physiological **activity** of an unknown chemical compound.

DERWENT CLASS: **B04** D16 **T01**
INVENTOR(S): CHAE, J H; SHIN, H C
PATENT ASSIGNEE(S): (SHIN-I) SHIN H C
COUNTRY COUNT: 1
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
KR 2002028925	A	20020417	(200363)*		1

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
KR 2002028925	A	KR 2002-990	20020108

PRIORITY APPLN. INFO: KR 2002-990 20020108

AB KR2002028925 A UPAB: 20031001
NOVELTY - A molecular design method is provided to construct a chromosome with weighting values of probes as genes, to optimize an average weighing value of the probes by applying a square method and a genetic **algorithm** alternatively or repeatedly, and analyzing a three dimensional quantitative **structure activity** relation so that it can estimate a physiological activity of an unknown chemical compound.

DETAILED DESCRIPTION - The method comprises steps of generating probes for calculating probe interaction energy (100), generating initial objects by expressing weighting values of the probes as genes (200), obtaining a linear coefficient of chromosome by using a square method for expressing a relation between the weighted probe interaction energy and the physiological activity (300), obtaining the average weighting value of the probes, and then obtaining the linear coefficient based the average weighting value (400), obtaining a better weighting value of the probe by using a genetic **algorithm**(600), and obtaining the physiological activity by using spatial coordinates, a partial charge, and a final weighting coefficient of the probes (700).
Dwg.1/10

L100 ANSWER 39 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-222583 [22] WPIDS
DOC. NO. NON-CPI: N2003-177411
TITLE: Expert system for planting management of alfalfa and preventing and eliminating diseases and pests.
DERWENT CLASS: P13 X25
INVENTOR(S): BAI, F; MA, Z; XIE, X
PATENT ASSIGNEE(S): (ZHIN-N) ZHINENGGU SCI & TECHNOLOGY CO LTD BEIJIN
COUNTRY COUNT: 1
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
CN 1381166	A	20021127	(200322)*		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
CN 1381166	A	CN 2002-121283	20020613

PRIORITY APPLN. INFO: CN 2002-121283 20020613
AB CN 1381166 A UPAB: 20030402
NOVELTY - An expert system for management and disease and pest prevention and elimination of alfalfa is composed of camera unit, intelligent controller, executing mechanism and planting the alfalfa in field. The growth state of alfalfa is picked up by camera and them compared with the management parameters in the **database**. After **inference** and analysis, the disease is judged and correct operating parametersare given out and are sent to the executing mechanism to apply related agricultural **chemical**. It can increase the yield of alfalfa by more than 10%.
Dwg.0/0

L100 ANSWER 40 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2002-195484 [25] WPIDS
DOC. NO. NON-CPI: N2002-148561
DOC. NO. CPI: C2002-060311
TITLE: Protein analysis in a biological system involves sampling the system after exposing it to a stimulus, treating the multiple samples by separation technique and analyzing the samples by parallel mass spectrometry.
DERWENT CLASS: B04 S03 V05
INVENTOR(S): JARDINE, I; LADINE, J R; STORY, M S
PATENT ASSIGNEE(S): (THER-N) THERMO FINNIGAN LLC; (JARD-I) JARDINE I; (LADI-I) LADINE J R; (STOR-I) STORY M S
COUNTRY COUNT: 96
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2001084143	A1	20011108	(200225)*	EN	43
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW					
AU 2001053462	A	20011112	(200225)		
US 2002068366	A1	20020606	(200241)		
EP 1274996	A1	20030115	(200306)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR					

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
WO 2001084143	A1		WO 2001-US12113	20010413
AU 2001053462	A		AU 2001-53462	20010413
US 2002068366	A1	Provisional	US 2000-196889P	20000413
			US 2001-835273	20010413
EP 1274996	A1		EP 2001-926964	20010413
			WO 2001-US12113	20010413

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2001053462	A	Based on	WO 2001084143
EP 1274996	A1	Based on	WO 2001084143

PRIORITY APPLN. INFO: US 2000-196889P 20000413; US 2001-835273 20010413

AB WO 200184143 A UPAB: 20020418

NOVELTY - Analysis of proteins in a biological system involves:

(a) exposing the system to a stimulus;

(b) sampling the system at multiple time intervals;

(c) treating the multiple samples by separation technique to provide multiple protein samples; and

(d) analyzing the multiple samples to determine changes in protein abundance as a function of time.

DETAILED DESCRIPTION - Analysis of proteins in a biological system involves:

(a) exposing the system to a stimulus;

(b) sampling the system at multiple time intervals;

(c) treating the multiple samples by separation technique to provide multiple protein samples; and

(d) analyzing the multiple samples to determine changes in protein abundance as a function of time.

The analysis includes directing mass spectral data from a parallel array of mass spectrometry systems to a common computing device and correlating the mass spectral data as a function of time. The mass spectral data is indicative of the identity and the abundance of protein in the multiple sample.

AN INDEPENDENT CLAIM is also included for a system for mass spectrometric analysis comprising:

(i) a parallel sample separation apparatus (A) adapted to separate multiple samples in parallel for analysis by mass spectrometry;

(ii) a parallel array of mass spectrometry systems (B) adapted to receive the samples from (A); and

(iii) a common computing device (C) communicating with (A) and (B). (C) is adapted to analyze the mass spectral data from (B) as a function of sample identity.

USE - For analyzing proteins in a biological system (claimed) e.g. a proteome, nucleotides or other **biological molecules**.

ADVANTAGE - The method achieves the analysis of a large number of proteins in an accurate, time-effective manner. The method allows to analyze the samples on a time scale governed only by the rate of the biological changes to observe and not by the rate at which the mass spectrometer performs the analysis. The method also allows one to **infer** the order of interactions between and among proteins without any advanced knowledge of pairs of interacting proteins as required by the protein interaction experiments. The potential for artifactual and false observation of protein interactions occurring in vitro is reduced as all protein interactions occur in vivo in their proper subcellular compartments. The method provides simultaneously recognition of multiple protein interaction pathways and their points of intersection. The method determines the time dependent appearance and disappearance of protein in normal cells compared to a cell treated with drug or perturbed by a disease or other factor. This is highly desirable in selecting alternative points of drug action in cases where the drugs have undesired reactions. The method not only increases and decreases in the abundance of particular proteins over time but also reveals shifts in structural state of those proteins with total abundance. The method identifies points at which protein modifications have occurred and reports the degree of modification of any protein.

Dwg.0/7

L100 ANSWER 41 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER: 2001-071535 [08] WPIDS

DOC. NO. NON-CPI: N2001-054116

DOC. NO. CPI: C2001-020116

TITLE: Methods, computer programs and **databases** for analyzing and make use of gene haplotype information.

DERWENT CLASS: B04 D16 S03 T01
INVENTOR(S): DENTON, R R; JUDSON, R S; RUANO, G; STEPHENS, J C; WINDEMUTH, A K; XU, C
PATENT ASSIGNEE(S): (GENA-N) GENAISSANCE PHARM INC
COUNTRY COUNT: 95
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2001001218	A2	20010104	(200108)*	EN	277
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TZ UG ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW					
AU 2000056386	A	20010131	(200124)		
EP 1208421	A2	20020529	(200243)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI					
EP 1233364	A2	20020821	(200262)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI					
EP 1233365	A2	20020821	(200262)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI					
EP 1233366	A2	20020821	(200262)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI					
JP 2003521024	W	20030708	(200347)		328

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
WO 2001001218	A2		WO 2000-US17540	20000626
AU 2000056386	A		AU 2000-56386	20000626
EP 1208421	A2		EP 2000-941722	20000626
			WO 2000-US17540	20000626
EP 1233364	A2	Div ex	EP 2000-941722	20000626
			EP 2002-7038	20000626
EP 1233365	A2	Div ex	EP 2000-941722	20000626
			EP 2002-7044	20000626
EP 1233366	A2	Div ex	EP 2000-941722	20000626
			EP 2002-7045	20000626
JP 2003521024	W		WO 2000-US17540	20000626
			JP 2001-507164	20000626

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2000056386	A	Based on	WO 2001001218
EP 1208421	A2	Based on	WO 2001001218
EP 1233364	A2	Div ex	EP 1208421
EP 1233365	A2	Div ex	EP 1208421
EP 1233366	A2	Div ex	EP 1208421
JP 2003521024	W	Based on	WO 2001001218

PRIORITY APPLN. INFO: US 1999-141521P 19990625
AB WO 200101218 A UPAB: 20011129
NOVELTY - Methods, computer programs and **databases** for analyzing and make use of gene haplotype (HT) information, e.g. to determine the

frequency of HTs in a population, to find correlations between HTs or genotype and a clinical outcome and to predict HTs from an genotype for a gene, are new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

(1) a method (M) for generating a HT **database** for a population, comprising data elements representative of the HTs for at least 1 locus from the individuals (indivs) in the data base;

(2) a (M) of predicting the presence of a HT pair in an indiv;

(3) a (M) for identifying a correlation between a HT pair and a clinical response to a treatment, or other phenotype Pt.;

(4) a (M) for identifying a correlation between a HT pair and a susceptibility to a condition or a disease of interest (OI), or other Pt. (OI);

(5) a (M) of predicting an indivs response to a medical or pharmaceutical treatment;

(6) a computer implemented (C-I) (M) for generating a gene structure screen for display on a display device (DD);

(7) a C-I (M) for generating a HT pair frequency screen for display on a DD;

(8) a C-I (M) for generating a linkage screen for display on a DD;

(9) a C-I (M) for generating a phylogenetic tree screen for display on a DD;

(10) a C-I (M) for generating a genotype (Gt.) analysis screen for display on a DD;

(11) a (M) of displaying clinical response values of a subject population as a function of HT pairs of the indivs in the population;

(12) a C-I (M) for carrying out a genetic **algorithm** for finding an optimal set of weights to fit a function of polymorphic site data to a clinical response measurement;

(13) a C-I (M) for displaying correlations between clinical outcome values for a selected population;

(14) a (M) for conducting a clinical trial of a treatment protocol for a medical condition (OI);

(15) a (M) of **inferring** Gts. of indiv subjects for a selected gene having polymorphic sites;

(16) a (M) of determining polymorphic sites or sub-HTs that correlate with a clinical response or out come (OI);

(17) a (M) of determining polymorphic sites or sub-HTs that correlate with a clinical response or outcome (OI);

(18) a computer usable (C-U) medium (Md.) having computer readable (C-R) program code (PC) stored upon it, for causing a computer (Comp.) to adjust observed HT pair frequencies within a population group (the HT pair frequencies are stored in a C-R **database** of HT information for a gene or gene feature (OI));

(19) a C-U Md. having C-R PC stored upon it, for causing HT pair assignments to be made to an indiv member of a population whose Gt. information for a gene feature (OI) is stored in a C-R form;

(20) a C-U Md. having C-R PC stored upon it, for causing a Comp. to identify a correlation between a clinical response to a treatment or other Pt. and a HT or HT pair present at a candidate locus associated with the clinical response or other Pt.;

(21) a C-U Md. having C-R PC stored upon it, for causing a Comp. to identify a correlation between an indiv's susceptibility to a condition or disease (OI) or other Pt., and a HT or HT pair present at a candidate locus associated with the susceptibility to the condition or disease (OI) other Pt. (OI);

(22) a C-U Md. having C-R PC stored upon it, for causing a Comp. to predict an indivs response to a medical or pharmaceutical treatment based on one or more selected HTs or HT pairs of the indiv;

(23) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display a gene's structure and gene features on a display device DD;

(24) a C-R Md. having C-R PC stored upon it, for causing a Comp. to

display on a DD, HT frequency data within a population of indivs, for a selected gene or gene feature;

(25) a C-R Md. having C-R PC stored upon it, for causing a Comp. to display on a DD, polymorphic site linkage data for a gene or gene (OI);

(26) a C-R Md. having C-R PC stored upon it, for causing a Comp. to display on a DD a phylogenetic tree;

(27) a C-R Md. having C-R PC stored upon it, for causing a Comp. to display a Gt. analysis screen on a DD;

(28) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display clinical response values, or other Pt. data, of a subject population as a function of HT pairs of the indivs in the population;

(29) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display on a DD, clinical response values, or other Pt. data, of a subject population as a function of HT pairs of the indivs in the population for a gene or gene feature (OI);

(30) a C-U Md. having C-R PC stored upon it, for causing a Comp. to carry out a genetic algorithm for finding an optimal set of weights to fit a function of polymorphic site data for a gene or gene feature (OI) to a clinical response measurement;

(31) a C-U Md. having C-R PC stored upon it, for causing a Comp. to display on a DD, correlation between clinical outcome values obtained from selected clinical outcome measures for a selected population;

(32) a C-U Md. having C-R PC stored upon it, for causing a Comp. to provide information of use in conducting clinical trials of a treatment protocol for a medical condition (OI);

(33) a C-U Md. having C-R PC stored upon it, for causing a Comp. to infer Gts. of indiv subjects for a selected gene having polymorphic sites;

(34) C-U media having C-R PC stored upon it, for causing a Comp. to determine polymorphic sites or sub-HTs that correlate with a clinical response or outcome (OI), or other Pt. (OI);

(35) Comps. programmed to carry out the above (Ms) or comprising the above Comp.-useable or -readable media, comprising a memory with at least 1 region for storing Comp. executable PCs and a processor for executing the PC stored in the memory;

(36) a data structure for storing an organizing biological information, stored on a C-R Md. and accessible by a processor, which comprises a single parent table which is adapted for storing, organizing and retrieving a number of genetic features by the relative positional relationships between the genetic features;

(37) a (M) for storing and organizing biological information; and

(38) a data structure for storing an organizing biological information, stored on a C-R Md. and accessible by a processor, which comprises a least 2 different fields, one of which included a number of genetic features, and the other of which included relative positional relationships between the genetic features.

Note: Further details of the above are given in the specification but had to be omitted from this abstract due to insufficient space.

USE - The methods, computer programs and databases for analyzing and make use of gene haplotype HT information, e.g. to determine the frequency of HTs in a population, to find correlations between HTs or genotypes and a clinical outcome or the effects of a therapeutic intervention and/or to predict HTs from an individual's genotype for a gene.

Dwg.0/49

L100 ANSWER 42 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2001-265296 [27] WPIDS
CROSS REFERENCE: 1998-159728 [14]
DOC. NO. NON-CPI: N2001-189700
DOC. NO. CPI: C2001-080181
TITLE: Generation of optimal quantitative **structure-activity** relationship among series of molecules, comprises using molecular hologram molecular structural **descriptor**.

DERWENT CLASS: **B04 T01**
INVENTOR(S): HERITAGE, T W; HURST, J R
PATENT ASSIGNEE(S): (TRIP-N) TRIPOS INC
COUNTRY COUNT: 1
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
US 6208942	B1	20010327	(200127)*		18

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
US 6208942	B1 CIP of	US 1996-698040	19960815
		US 1998-22252	19980210

FILING DETAILS:

PATENT NO	KIND	PATENT NO
US 6208942	B1 CIP of	US 5751605

PRIORITY APPLN. INFO: US 1998-22252 19980210; US 1996-698040 19960815

AB US 6208942 B UPAB: 20010518
NOVELTY - Generation of optimal quantitative **structure-activity** relationship (QSAR) among a series of molecules, comprising using a molecular hologram molecular structural descriptor for each molecule in the series, where each molecule is associated with an activity value, is new.

DETAILED DESCRIPTION - Generation of optimal quantitative **structure-activity** relationship among a series of molecules comprising:

(a) defining a list of values for hologram length and fragment size range;

(b) selecting a value from the list for length L;

(c) selecting values from the list for fragment size in M-N;

(d) using selected values of M and N which define a molecular hologram molecular structural descriptor for each molecule in the series;

(e) correlating the molecular hologram molecular **structural** descriptor and **activity** value of each molecule with all the other molecules to obtain a **structure-activity** relationship;

(f) repeating steps (b)-(e) for all values of L on the list;

(g) selecting the optimal **structural-activity** relationship based on the statistical correlation values; and

(h) outputting the selected optimal **structure-activity** relationship for the values of L-N used for the molecular hologram generation along with statistical significance measurements

An INDEPENDENT CLAIM is also included for generating a weighted 2-Dimensional (2D) fingerprint of a molecule comprising:

(1) generating a list of all fragments found in the molecule having a minimum size of M and maximum size of N including branched and cyclic fragments;

(2) producing a unique representation of each fragment;

(3) generating each unique representation of each pseudo-random number generated by fragment;

(4) assigning each fragment to a specific position in the fingerprint using operator modulus with the length L and the pseudo-random number; and

(5) incrementing the value stored at each fragment position for each occurrence in the molecule assigned to that position.

USE - For generating optimal quantitative **structure-**

activity relationship among series of molecules.
ADVANTAGE - Powerful chemometric techniques are applied to the molecular holograms to yield predictive quantitative **structure-activity** models. The process determines the optimal set of parameters to use in hologram generation so that the resultant hologram yields the optimal validated QSAR model. It provides huge benefits to the user and extends the scope of quantitative **structure-activity** relationship (QSAR) modeling to a wider application, e.g. CoMFA or Apex-3D. The technique can be automated.
Dwg.0/9

L100 ANSWER 43 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2000-665269 [64] WPIDS
DOC. NO. NON-CPI: N2000-493033
DOC. NO. CPI: C2000-201590
TITLE: Identifying novel nucleic acid molecules encoding proteins of interest, and natural language processing and extraction of **relational** information associated with genes and proteins found in journal articles.
DERWENT CLASS: B04 D16 S03 T01
INVENTOR(S): FRIEDMAN, C; KALACHIKOV, S; KRA, P; KRAUTHAMMER, M O; RZHETSKY, A
PATENT ASSIGNEE(S): (UYCO) UNIV COLUMBIA NEW YORK; (KALA-I) KALACHIKOV S; (RZHE-I) RZHETSKY A
COUNTRY COUNT: 92
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2000063687	A1	20001026	(200064)*	EN	374
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW					
AU 2000043556	A	20001102	(200107)		
US 2002049542	A1	20020425	(200233)		
US 6633819	B2	20031014	(200368)		

APPLICATION DETAILS:

PATENT NO	KIND		APPLICATION	DATE
WO 2000063687	A1		WO 2000-US10302	20000414
AU 2000043556	A		AU 2000-43556	20000414
US 2002049542	A1	Provisional	US 1999-129469P	19990415
			US 1999-327983	19990608
US 6633819	B2	Provisional	US 1999-129469P	19990415
			US 1999-327983	19990608

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 2000043556	A	Based on	WO 2000063687

PRIORITY APPLN. INFO: US 1999-327983 19990608; US 1999-129469P 19990415
AB WO 200063687 A UPAB: 20001209
NOVELTY - Identifying novel nucleic acids molecules encoding a protein of **interest**, using regulatory networks, is new.
DETAILED DESCRIPTION - Identifying novel nucleic acids molecules

encoding a protein of interest, using regulatory networks, is new. The method comprises:

(a) selecting a specific protein from a species involved in a regulatory network of interest;

(b) identifying known proteins that act upstream and downstream of the protein, within the regulatory network;

(c) constructing the regulatory network of interest from the proteins identified in (b);

(d) for each identified protein, selecting a domain or motif and searching by homology for related proteins in a second species, a related protein has a homologous domain or motif;

(e) producing a regulatory network for the second species, which incorporates the identified related proteins;

(f) comparing the networks of the two species;

(g) identifying a protein present in only one of the networks; and

(h) isolating a nucleic acid molecule encoding the protein identified in (g) in the species in which it is missing.

INDEPENDENT CLAIMS are also included for the following:

(1) identifying the effect of a gene knockout on a regulatory pathway, comprising:

(a) identifying the shortest non-oriented pathway connecting two gene products;

(b) assigning an initial sign value of minus to the knockout since the knockout gene is inactive;

(c) moving along the shortest pathway between the two gene products multiplying the sign with the sign of the next gene product in the pathway, where minus stands for inhibition and plus stands for induction or activation and zero stands for lack of interaction between two proteins in the specified direction; and

(d) determining the final sign at the end of the pathway, where minus indicates inhibition and plus indicates induction or activation of the pathway;

(2) identifying a novel nucleic acid molecule encoding a protein of interest, comprising:

(a) selecting a gene of interest and searching a **database** for homologous sequences;

(b) aligning the sequences identified in (a);

(c) constructing a gene tree using the sequence alignment;

(d) constructing a species tree;

(e) inputting the species tree and gene tree into an **algorithm** which integrates the species tree and gene tree into a reconciled tree; and

(f) identifying orthologous genes present in one species but missing in another;

(3) identifying a novel gene, comprising:

(a) defining a motif or domain composition of a gene of interest;

(b) searching for sequences which correspond to nucleotide sequences in an expression sequence tag **database** or other cDNA **database** using a program such as BLAST and retrieving the identified sequences;

(c) searching additional **databases** for expressed sequence tags containing the domains and motifs characteristic for the gene of interest with a hidden Markov model of domains and motifs identified in (A); and

(d) identifying nucleotide sequences comprising the gene of interest;

(4) extracting information on interactions between biological entities from natural-language text data, comprising:

(a) parsing the text data to determine its grammatical structure; and

(b) regularizing the parsed text data to form structured word terms; and

(5) a computer system for extracting information on biological entities from natural-language text data, comprising:

(a) means for parsing the natural-language text data; and

(b) means for regularizing the parsed text data to form structured word terms.

USE - For identifying novel genes and for natural language processing and extraction of relational information associated with genes and proteins that are found in genomics journal articles.

ADVANTAGE - The method allows the rapid retrieval of information from literature and manipulation of derived functional data, removing a researchers need to perform laborious reading and manual integration of research articles.

Dwg.0/23

L100 ANSWER 44 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2000-271477 [23] WPIDS
DOC. NO. CPI: C2000-082969
TITLE: Automated discovery of genomic data, useful for developing e.g. drugs or pesticides, by parallel, iterative knowledge discovery for many genes in a database.
DERWENT CLASS: B04 D16
INVENTOR(S): CARIASO, M C; SHI, Q; STEWARD, K L
PATENT ASSIGNEE(S): (GENE-N) GENE LOGIC INC
COUNTRY COUNT: 88
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2000015847	A2	20000323	(200023)*	EN	54
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ UG ZW					
W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZA ZW					
AU 9962440	A	20000403	(200034)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2000015847	A2	WO 1999-US20449	19990908
AU 9962440	A	AU 1999-62440	19990908

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9962440	A	Based on	WO 2000015847

PRIORITY APPLN. INFO: US 1998-100030P 19980911

AB WO 200015847 A UPAB: 20000516

NOVELTY - A method for genomic data discovery, comprises discovering knowledge, in parallel, about all selected genes in a database of at least 10 genes (I), and using the acquired knowledge to repeat the procedure several times.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) a method for discovering genomic knowledge, comprising determining at least one data element (DE), for at least one (I), searching at least 50 databases for this DE, and analyzing responses to increase knowledge of (I);

(2) a method of automated knowledge discovery comprising continuously operating a cycle that comprises querying a database to receive data, drawing inferences from the data to generate knowledge, and re-evaluation

of the **inferences** when the **database** is modified;

(3) a method for discovering genomic knowledge by selecting a gene token (GT), determining data requirements for GT, requesting and receiving data responsive to these requirements, analyzing the information to increase knowledge of GT and repeating the procedure at least 50 times;

(4) a knowledge discovery system comprising a unit for determining data needs and analyzing returned data responsive to these needs, and at least 10 adapter units for accessing at least 10 dissimilar data sources to provide the data required;

(5) a method of ranking (I) for a particular application by computer-based application, without additional operator input, of application-specific ranking rules to many GT;

(6) a method of genomic information analysis, comprising applying inference rules to two models of a **biological** relationship, interrelating different sets of genes or proteins, and applying inference rules to the models to infer missing information; and

(7) an automated method of genomic knowledge discovery by analyzing GT to determine required data and either, asking a human expert for data or generating by computer, without additional operator input, a work order to a laboratory to produce the data.

USE - The method is used for the development of drugs, cosmetics, food additives, pesticides, herbicides and other biologically active agents. More generally similar methods can be used to process industrial or financial information.

ADVANTAGE - The method is automated to allow manipulation of more information than could be handled by a human operator, i.e. it overcomes difficulties associated with scale, updating, errors, heterogeneity and complexity of databases. It can be operated continuously to take account of changes in knowledge and/or available resources, both external and internal, and may include self-monitoring to identify the most dependable data sources or to identify/correct errors.

Dwg.0/5

L100 ANSWER 45 OF 45 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 1997-244768 [22] WPIDS
DOC. NO. NON-CPI: N1997-201912
DOC. NO. CPI: C1997-079264
TITLE: Preparing **database** of molecular fragments by counting all fragments in a molecule - and storing counts in computer memory, useful for analysing **structure-activity** relationships, especially of drugs and toxins.
DERWENT CLASS: **B04** J04 S03 **T01**
INVENTOR(S): BONE, R G A; VILLAR, H O
PATENT ASSIGNEE(S): (TERR-N) TERRAPIN TECHNOLOGIES INC
COUNTRY COUNT: 20
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9714106	A1	19970417	(199722)*	EN	48
RW: AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE					
W: AU CA JP					
AU 9673987	A	19970430	(199734)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9714106	A1	WO 1996-US16196	19961010
AU 9673987	A	AU 1996-73987	19961010

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9673987	A	Based on	WO 9714106

PRIORITY APPLN. INFO: US 1995-550847 19951031; US 1995-542642 19951013

AB WO 9714106 A UPAB: 19970530

Database of molecular fragments is prepared by:

(a) identifying all sequentially attached fragments within a selected molecule,

(b) counting the occurrences of each unique fragment and

(c) storing information correlating fragment counts with fragment identity in computer-readable form.

Also claimed are:

(1) data processing system for creating such **databases**, and

(2) computer-readable medium on which the **databases** are stored.

USE - Comparison of fragment counts between a molecule and a reference molecule of known activity can be used to predict which compound will have this particular activity. The method is especially applied to libraries of drugs (e.g. central nervous system drugs), toxins or randomly chosen compounds.

ADVANTAGE - The **databases** provide a complete and systematic classification of function/activity based on specific topological characteristics of fragments of small molecules, and is suitable for construction of combinatorial libraries covering the whole of chemical space or focused on part of it for precise selection of active molecules. Dwg.13a/18

FILE 'HOME' ENTERED AT 16:04:42 ON 22 OCT 2003